The Effect of Context on the Activation and Processing of Word Meaning Over Time

Diego Frassinelli

Doctor of Philosophy

Institute for Language, Cognition and Computation
School of Informatics
University of Edinburgh

2015

Abstract

The aim of this thesis is to study the effect that linguistic context exerts on the activation and processing of word meaning over time. Previous studies have demonstrated that a biasing context makes it possible to predict upcoming words. The context causes the pre-activation of expected words and facilitates their processing when they are encountered. The interaction of context and word meaning can be described in terms of feature overlap: as the context unfolds, the semantic features of the processed words are activated and words that match those features are pre-activated and thus processed more quickly when encountered. The aim of the experiments in this thesis is to test a key prediction of this account, viz., that the facilitation effect is additive and occurs together with the unfolding context.

Our first contribution is to analyse the effect of an increasing amount of biasing context on the pre-activation of the meaning of a critical word. In a self-paced reading study, we investigate the amount of biasing information required to boost word processing: at least two biasing words are required to significantly reduce the time to read the critical word. In a complementary visual world experiment we study the effect of context as it unfolds over time. We identify a ceiling effect after the first biasing word: when the expected word has been pre-activated, an increasing amount of context does not produce any additional significant facilitation effect.

Our second contribution is to model the activation effect observed in the previous experiments using a bag-of-words distributional semantic model. The similarity scores generated by the model significantly correlate with the association scores produced by humans. When we use point-wise multiplication to combine contextual word vectors, the model provides a computational implementation of feature overlap theory, successfully predicting reading times.

Our third contribution is to analyse the effect of context on semantically similar words. In another visual world experiment, we show that words that are semantically similar generate similar eye-movements towards a related object depicted on the screen. A coherent context pre-activates the critical word and therefore increases the expectations towards it. This experiment also tested the cognitive validity of a distributional model of semantics by using this model to generate the critical words for the experimental materials used.

Lay Summary

Several studies have highlighted the key role of contextual information in facilitating the processing of upcoming words. Facilitation has been motivated in terms of feature overlap: as the context unfolds, a certain number of semantic features become active. The words that match those features are pre-activated and consequently processed more quickly.

This study contributes to the debate, with a fundamental element of novelty: the manipulation of the amount of the biasing contextual information provided. Controlling the nature and the amount of this information allows us to directly analyse the interaction between the context and the target word. We present results from self paced-reading and visual world experiments as well as corpus-based modelling of behavioural data with a co-occurrence distributional model.

Taken together, the results show that a biasing context affects word processing. This facilitation process is cumulative but it requires a certain amount of contextual information in order to become active. The distributional semantic model (as point-wise vector multiplication) can capture the feature overlap theory, successfully predicting reading times.

Acknowledgements

The first big thank you goes to Frank Keller, a good and patient supervisor who taught me how to be always honest and systematic in research. This work would not have been possible without his guidance, support and encouragement. A thanks also to Christoph Scheepers for his feedback, suggestions and new ideas.

I am grateful to Ken McRae and Patrick Sturt for making the Viva such a stimulating experience.

Thanks to all the ILCC crowd that made my time in Edinburgh memorable: Aciel Eshky, Dave Matthews, Dominika Lyzwa, Greg Coppola, Luke Shrimpton, Mike Lewis, Sasa Petrovic and many others. I am also grateful to the level 3 admin people for their daily help. Thanks to Caroline Domenech and Lukas Michelbacher, an amazing last minute acquisition.

I am extremely thankful to Desmond Elliott, Eva Hasler, Silvia Pareti and Stella Frank; you have been always there when I needed help, support or just a "good" coffee or a nice IPA.

Special thanks to Alessandra Zarcone, great friend and researcher; Moreno Coco, for all the time he spent trying to introduce me to the magic world of statistics; Raffaele Limosani, the old good friend; and Gabriella Lapesa who, after so many years, was still there even in the last crazy moments before the submission.

In these years, I had amazing flatmates: a big family always up for a chat or a dinner and ready to provide unconditional support when needed: Alessia Tosi, Andrew McFarlane, Diana Bankovic, Erman Ozgur, Lena Moller, and Nese Karahasan. Thank you so much guys, you definitely made my Edinburgh.

Thanks again to Andrew, Des, Eva, Gabriella, and Stella for their immense feedback to this thesis.

Last but not least, grazie a mamma e a Sergio per essere stati presenti e per aver sempre sostenuto il mio continuo vagabondare senza fare troppe domande.

This research was supported by the EPSRC and by the Scottish Informatics and Computer Science Alliance (SICSA).

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Diego Frassinelli*)

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The incremental nature of language directly affects the way humans process linguistic information. It has been argued that language is incremental and contextual information becomes available over time (Altmann & Steedman, 1988). Contextual information plays an essential role in language processing; it has been demonstrated that a coherent linguistic context makes the processing of the upcoming information in the sentence faster and more accurate (Kutas & Federmeier, 2000).

The interaction of the words that occur together in a sentence takes place at different levels of interpretation: e.g. prosody, syntax, semantics, and pragmatics. In the literature several studies have been conducted to analyse these different levels individually. In particular, the studies focusing on the semantic dimension of this interaction show that the linguistic context strongly influences the activation and processing of word meaning (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007).

In the psycholinguistic literature there has been wide discussion on the exact moment when context affects word processing. The use of eye-tracking and neurophysiological techniques have shown that when processing an unfolding stream of linguistic information, we actively update the expectations of upcoming words (Kutas & Hillyard, 1984). This evidence is in favour of the *predictive account*: context preactivates some of the most expected meanings and facilitates the processing of the corresponding words when encountered. The predictive power of context has been described in terms of *feature overlap* (Federmeier & Kutas, 1999): as the context unfolds, it activates semantic features that restrict the set of possible upcoming words. The words that match these features are pre-activated and thus processed more quickly when encountered.

A key assumption of the predictive account is that the facilitation effect is additive and occurs together with the unfolding context. In the experiments reported in this thesis we directly test this assumption manipulating the amount of biasing words in the context.

1.1 Thesis Overview

In order to analyse the interaction between context and word meaning we propose two possible hypotheses that are both compatible with the predictive account. On the one hand, the *incremental activation hypothesis* directly reflects the main assumptions of this account: the facilitation effect starts with the first context word provided and increases over time with the amount of available context. On the other hand, the *immediate activation hypothesis* requires a certain number of highly related context to boost the pre-activation of the critical word. Moreover, this second hypothesis predicts the existence of a ceiling effect: after a certain amount of highly related context has been provided there is no more additional significant facilitation taking place. The experiments performed in this thesis confirm the predictive nature of context: it allows the pre-activation of the meanings of the upcoming words. The results obtained support the validity of the *immediate activation hypothesis* that imposes some constraints to the facilitation process. The same effect can be described in terms of *feature overlap*.

In Chapter 2 we review the most important studies on the effect of context in language processing. We report the big debate that concerns when and how context affects the activation of word meaning. Moreover, we discuss how sentence constraints and semantic relations facilitate the processing of the upcoming words. Finally, we show how contextual effects can be described in terms of feature overlap.

In Chapter 3 we provide background information about the representation of word meaning as semantic properties. We present distributional semantic models as a way to automatically extract semantic information from linguistic corpora. We also report a study that tests the cognitive validity of distributional semantic models in capturing semantic relations between words.

In Chapter 4 we describe the methodologies we used to conduct the experiments reported in the thesis: the self-paced reading paradigm, the visual world paradigm and the blank screen paradigm.

In Chapter 5 we investigate the effect of increasing the amount of biasing information on word processing. The experimental manipulation allows us to distinguish between the incremental and the immediate activation hypotheses. In order to describe the interaction between context and word meaning we perform four experiments. The linguistic stimuli used in these experiments contain context words that are semantically more related to the final critical word (high biasing words) and context words that are less related to it (low biasing words). In a self-paced reading experiment (**Experiment 1**) we test the effect of an increasing amount of biasing information on the reading time of the critical word. Given that this word occurs always at the end of the sentence (followed by some spill-over material) we can successfully study the cumulative effect of context. The outcome of Experiment 1 indicates that at least two biasing words are required to pre-activate the meaning of the critical word. The inclusion of more biasing information does not produce any significant effect.

In order to introduce the time course in our analysis, we perform a visual world experiment (**Experiment 2**). This experimental paradigm measures the activation of word meaning driven by context in terms of the amount of fixations towards a related object depicted on the screen. We use this technique to study the activation effect produced by each word in the context. The results of Experiment 2 indicate that one biasing context word is enough to pre-activate the meaning of the critical word. Experiment 1 and Experiment 2 support the immediate activation hypothesis. They demonstrate a facilitation effect driven by context: a certain amount of biasing information is required to boost word processing and a ceiling effect occurs when the pre-activation has taken place.

The stimuli of the previous experiments followed a predetermined order of context words (always a location, followed by an actor and finally an object). We perform a second self-paced reading study (**Experiment 3**) in order to understand the relationship between order and the semantic nature of context words. For this study we designed a new set of linguistic stimuli which manipulate the order of the context words in the sentence (e.g. having an actor or an object as the first context word). These stimuli are composed by a main clause containing the three context words and a subordinate clause with the critical word. The statistical differences that emerged in the previous studies were not replicated in this experiment. The results suggest that encapsulating the critical word in a subordinate clause isolates it from the influence of the context.

Once we have determined the effect of the entire contextual sentence on the processing of the critical word, we investigate the relation between each context word and the critical one as well as between the context words themselves. In four association studies (**Experiment 4**) we report the semantic relations between those words. Overall, we show that context words that are highly related to the critical word are also highly associated to each other.

Having analysed how context words bias the pre-activation of the critical word, we describe the same biasing effect in terms of feature overlap in Chapter 6. Distributional semantic models represent the meaning of a word as a multidimensional vector: every dimension of the vector corresponds to another word appearing in its contextual proximity (Sahlgren, 2006). It is possible to consider each dimension as a semantic feature contributing to the overall meaning of the word described. For this reason, a distributional model is a good candidate to describe the feature overlap theory. A series of correlation studies show that the similarity scores generated by a distributional semantic model significantly correlate with the association scores produced by humans in Experiment 4. Further, we demonstrate that it is possible to model the feature overlap theory in terms of vector combination. When we combine the contextual word vectors by point-wise multiplication the model successfully predicts the reading times of Experiment 1.

Finally, in Chapter 7 we provide further evidence in favour of the feature overlap theory by analysing how semantic similarity facilitates the access of word meaning. Two words are similar if they share semantic features. For this reason, they should generate similar effects in word processing. In a visual world experiment (**Experiment 5**) we analyse how the critical word guides eye-movements towards the picture of the object associated with that word or with a semantically similar word. We also manipulate the semantic similarity relation between context words and the critical word. We design the linguistic stimuli using three words generated by a distributional semantic model. We show that semantically related context allows the pre-activation of the meaning of the critical word, increasing the expectations towards it. We also test the cognitive validity of distributional semantic models in generating the stimuli for this study. These words can be successfully embedded into experimental stimuli and guide eye-movements towards a related object depicted on the screen.

In a blank screen paradigm study (**Experiment 6**) we partially replicate the outcomes of Experiment 5. The aim of this experiment was to reduce the effect of the visual stimulus while processing linguistic information. Overall, the absence of the visual stimulus produces a significant reduction in the amount of fixations in the direction of the area of the screen where the target object was depicted. However, the existence of an expectation produced by contextual information can be observed also when the visual scene is not present.

1.2 Main Contributions

The general aim of this thesis is to investigate the effect of context on the activation and processing of word meaning over time. We want to test the assumption behind the predictive account: we ask when context starts affecting word processing and whether this effect is additive.

The first, and most important contribution of this work, is the analysis of how word processing is affected by increasing amounts of biasing context. Experiment 1 and Experiment 2 show evidence in favour of the immediate activation hypothesis: context pre-activates the meaning of the critical word. Based on the methodology adopted to perform the experiment, one or two biasing words are required to significantly boost pre-activation. Once pre-activation has taken place, an increasing number of biasing words does not produce any additional significant effect.

The second contribution involves the description of the feature overlap theory in terms of distributional semantics. We show that a bag-of-words distributional semantic model can successfully represent the cumulative facilitation on word activation and processing driven by context. We demonstrate that the accumulation of semantic features can be modelled as the composition of the distributional vectors of the context words. Distributional semantics therefore provides a computational implementation of feature overlap theory, with semantic features represented as vector components (i.e., word co-occurrences).

The third contribution regards the analysis of the effect of context on semantically similar words. We show that words that are semantically similar generate similar eye-movement patterns. We also show that the semantic relation between context and target words facilitates the processing of the critical word. The outcome of these experiments adds evidence to the idea that pre-activation of word meaning is driven by context at the property level. In the same experiment we also tested the cognitive validity of a distributional model of semantics. We used the words extracted by the model to produce the stimuli for these experiments.

1.3 Publications

The self-paced reading (Experiment 1) and the visual world paradigm (Experiment 2) experiments described in Chapter 5 are published in Frassinelli, Keller, and Scheepers (2013). The visual world experiment (Experiment 5) described in Chapter 7 appears in Frassinelli and Keller (2012).

Chapter 2

From Perception to Meaning: Contextual Effects on Word Processing

In this chapter we review the most important studies on contextual effects in language processing. Context is a very rich source of information and it has been recognised to strongly facilitate different cognitive tasks: e.g. visual processing (Oliva & Torralba, 2007; Torralba, Oliva, Castelhano, & Henderson, 2006; Hwang, Wang, & Pomplun, 2009), and language processing. In the language processing literature, several studies aim to answer the question "how does the human brain so effectively move from perception to meaning?" (Federmeier, 2007, p.491). Various analyses have been performed in order to understand the role played by contextual information. The first studies were priming experiments manipulating the relation between one or more words and the target one. Those studies mainly included reaction times (e.g. word / non word discrimination) and reading times. The introduction of eye-tracking and neurophysiological approaches in the study of language processing has allowed for the development of a more precise on-line analysis over time. Overall, studies performed in the last forty years agree on the importance of context in language processing. The effect of context on word processing can be driven by different linguistic and extra-linguistic elements that have been individually analysed. A facilitator effect can be related to phonological elements (e.g. the phonemic restoration effect in Marslen-Wilson & Welsh, 1978), semantic and syntactic relations between context and target words (e.g thematic role assignment in Altmann, 1999).

In the following sections we review the main studies that have been conducted on semantic processing at word level. We are mainly interested in understanding when and how this process takes place. After discussing the importance of context in language comprehension (Section 2.1), in Section 2.2 we describe the two main accounts explaining when contextual information plays its role. After that, in Section 2.3, we report the main studies manipulating the constraint exerted by context. In Section 2.4 we discuss the effect of context in terms of semantic feature overlap.

2.1 Introduction

Human beings are usually very good at completing a sentence with the most plausible word. For example, when asked to complete the following sentences:

(1) The kid was on the sofa in front of the ...

(2) The kid was on the sofa watching the ...

the majority of us would easily use the word "television" in both cases. In Example (1), television is the most plausible completion even though it does not have a strong lexical or semantic association with the sentence words. The task becomes even easier for Example (2) in which a related word is provided (e.g. "watching").

The semantic context surrounding a specific word plays different roles in the interpretation of that word. For example, in Miller (1978) the interpretation of the adjective "good" changes based on the noun it is modifying: "a good meal", "a good deal", or "a good book". In Anderson et al. (1976), a general noun obtains specific interpretations based on the context: in the sentence "the container held the cola" the container is interpreted as a "bottle"; while in the sentence "the container held the apples" the container is recognised as a "basket". In Johnson-Laird (1987), different contexts highlight different properties of a word (semantic flexibility). For example, in the sentences "the tomato rolled across the floor", "the sun was a ripe tomato", and "he accidentally sat on the tomato", different features of the tomato are highlighted: the shape, the colour, and the texture (the examples for the three studies are taken from Moss & Marslen-Wilson, 1993, p.1254).

As is already clear from the examples above, context plays two main roles: it facilitates the identification of (one of) the most plausible completions (inhibiting the implausible counterparts) and it guides the interpretation of the upcoming information. A substantial body of research has demonstrated the facilitatory effect of a congruent and strongly biasing context: words that are more predictable are also accessed and processed faster. Naming studies (Duffy, Henderson, & Morris, 1989; McClelland & O'Regan, 1981; Stanovich & West, 1983) and lexical decision studies (Fischler & Bloom, 1979; Jordan & Thomas, 2002) show an increased precision and speed when contextual information is strongly constraining. Similarly, reading studies show that strongly constrained words are often skipped or fixated for less time than weakly constrained words (Ehrlich & Rayner, 1981; McDonald & Shillcock, 2003; Morris, 1994). Moreover, ERP studies (Kutas & Hillyard, 1980; see Kutas & Federmeier, 2000 for an extensive review) report a smaller N400 amplitude when the word appears in a supportive context. These studies considered mainly semantic relations. However, there is a more recent trend towards considering also syntactic effects. For example, the thematic fit literature focuses on the probability of a noun to be the argument of a specific verb (Bicknell, Elman, Hare, McRae, & Kutas, 2010; Kamide, Altmann, & Haywood, 2003; McRae, Spivey-Knowlton, & Tanenhaus, 1998). The studies on syntactic effects showed that the event associated with a specific verb affects the processing of the upcoming information.

All these studies agree on the key role of contextual information in language processing. However, the nature and strength of the interaction between context and upcoming words has been a source of intense debate (Van Petten & Luka, 2012). The most discussed point regards the moment when context exerts its influence. The behavioural literature tends to favour the post-lexical account. While Sedivy, Tanenhaus, and Chambers (1999) suggest an incremental processing that produces an early integration of contextual information. This view was already anticipated in Altmann and Steedman (1988) and agrees with the eye-tracking (Altmann & Kamide, 1999) and ERP literature (Kutas & Hillyard, 1980). Both these techniques allow a more precise study over time and show an early integration of contextual effects.

2.2 Language Comprehension: Integration or Prediction?

In the psycholinguistic literature, a big debate concerns when and how context affects the lexical access to long term memory: immediately after becoming available or only post-lexically, when the word is already activated. A big contribution to understanding this problem comes from the ambiguity resolution literature (see Swaab, Brown, &

Hagoort, 2003 for an extensive discussion). For example, consider the following two sentences related to two different meanings of the polysemous word "bank":

(3) The businessman withdrew all his money from the *bank*

(4) The fisherman was paddling his boat toward the *bank*

Sentence (3) refers to the bank as the financial establishment and it is also the preferred (most frequent) meaning, while Sentence (4) refers to the bank as the land alongside the river.

For many years, the *integrative account* (also called post-lexical access, context-independent multiple access, or exhaustive access in different sub-disciplines) was considered the "standard position" (Williams, 1988) for explaining the effect of context on word recognition and processing (Conrad, 1974; Lucas, 1987; Swinney, Onifer, Prather, & Hirshkowitz, 1979). According to the modular perspective of this account (Fodor, 1983), only after accessing the lexical and semantic information of a specific word, can previous contextual information be successfully integrated. The access to word meaning is completely data-driven (bottom-up) and takes place in a "window of contextual impenetrability" (Prather & Swinney, 1988, p.290). For example, both in Sentence (3) and in Sentence (4) the two different meanings of bank would be equally pre-activated. Only when the activation has taken place, contextual information would allow the selection of the most plausible meaning. Results corroborating the integrative account appear in lexical decision tasks (Tanenhaus, Leiman, & Seidenberg, 1979) and naming studies (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Van Petten & Kutas, 1988). A strong evidence in favour of this account comes from the cross-modal lexical priming literature (Swinney et al., 1979). This paradigm integrates the effects of different sensory modalities (auditory and visual) in a lexical decision task. For each trial, the first linguistic item is presented auditory and it is immediately followed by the second linguistic item presented visually. Participants have to decide whether the first auditory item is a legal word or not. If the answer is positive, they have to perform the same judgement for the second visual item. The reaction times are recorded. Often the results of these experiments provide the evidence that during a lexical decision task participants access all the meanings of a word independently by the context provided (Onifer & Swinney, 1981).

Similarly, the *ordered access account* indicates that the most frequent meaning of a word is always retrieved and activated first, while less frequent meanings are activated only when clearly anticipated by context (Hogaboam & Perfetti, 1975).

On the other hand, the *predictive account* (also called selective access, or context-dependent access) suggests that contextual information is used immediately after becoming available and affects lexical access continuously over time. When encountering a word, only the most appropriate meaning (already predicted) is activated and processed: for example, the financial establishment in Sentence (3) and the land alongside the river in Sentence (4). This account recognises the existence of a top-down process that treats the linguistic stimulus based on previous contextual knowledge (McClelland & Elman, 1986; McClelland & Rumelhart, 1981). Over time, the meanings of upcoming words are evaluated against the predictions and constraints imposed by the general context (Moss, 1997). Similarly, the *constraint satisfaction account* proposes that the identification of the most plausible meaning is driven by both lexical and contextual factors: a strong lexical bias (e.g. high word frequency) would overcome the effects of the context (MacDonald, Pearlmutter, & Seidenberg, 1994).

Tabossi and colleagues support the predictive account describing a facilitation effect driven by context both in ambiguous (Tabossi, Colombo, & Job, 1987) and unambiguous words (Tabossi, 1988). They use the same cross-modal priming paradigm that Swinney et al. (1979) used to support the integrative account. The experiments performed show that if there is a strong association (in terms of semantic information constraints) between the context and one meaning of the ambiguous word, this meaning is activated faster. In the same study the authors reveal a strong interaction of contextual bias and lexical effects (mainly frequency) in the facilitation process. This evidence suggests that, when context refers to the most frequent meaning of a word, the other meanings are not activated at all. Similarly, Rayner and Frazier (1989) show that frequency guides the access to word meaning when the contextual constraints are not strong enough. However, if there is a strong association between contextual words and one meaning of an ambiguous word, this meaning will be activated faster (Traxler & Foss, 2000).

More recently, Jackendoff (2002) considered the predictive account as cognitively implausible for the following reason. Often contextual information is not strong enough to successfully guide the activation of a specific meaning. Dealing with the prediction of the wrong meaning would be extremely expensive from a cognitive perspective.

A possible mediation between the two accounts is proposed by Federmeier (2007). The author suggests that the complexity of the task requires multiple parallel processes taking place in both brain hemispheres. Prediction occurs mainly in the left hemisphere of the brain, leaving to the right hemisphere the task of checking for the actual accuracy of the prediction, and in the case of error, to initiate a correction process.

2.3 Modulating Sentence Constraints

Together with word frequency, contextual constraints play the main role in the facilitation of language processing. The cloze probability paradigm is a traditional way of quantifying the relation between contextual information and a specific word (the target) in a sentence (Taylor, 1953). Participants are asked to complete a sentence with the word that best fits into it. The cloze probability is evaluated as the number of participants that completed the sentence using a specific word out of the total number of participants. This measure is extensively used to quantify the biasing effect of context. A highly constraining context produces a more precise facilitation for the most expected words than lowly constraining contexts (Fischler & Bloom, 1979; McClelland & O'Regan, 1981). For example, Ehrlich and Rayner (1981) studied the effect of context in guiding eye-movements in a reading task. They reported an inverse relation between contextual constraints and fixation durations: the stronger the contextual constraints, the shorter the fixation time toward the expected word. Similarly, Morris (1994) studied eye-movements towards unambiguous nouns in sentence contexts. Shorter reading-times are related to stronger contextual associations, even when the syntactic relation between words changes.

Schwanenflugel and LaCount (1988) analysed the effects of high and low constraining sentences in facilitating the processing of expected and unexpected words (see also Schwanenflugel & Shoben, 1985). The authors discussed the combined effect of sentence constraints and semantic relatedness as factors that facilitate the processing of the final word of a sentence. Overall, these studies show that high-constraint sentences facilitate only the most expected completions, while low-constraint sentences produce a wider but also weaker facilitation effect on the upcoming words. They discuss this effect in terms of *featural restrictions* that participants generate when they are exposed to contextual information. When fewer restrictions are imposed by the context, more words can successfully fill the final position. On the other hand, when more restrictions are provided, less words can fill the specific slot. To illustrate this effect, the authors use the following examples:

(5) Hank reached into his pocket to get the ...

(6) The tired mother gave her dirty child a ...

Sentence (5) presents a low-constraint sentence. In this case, the possible featural restrictions generated are: [frequently found in pockets] and [small]. A higher amount of possible completions can occur in the final slot given the small amount of constraints. On the other hand, Sentence (6) shows a high-constraint sentence where the amount of restrictions is higher and more specific: e.g. [cleans], [common to children], [given by mothers], [taken by people]. The authors discussed featural restriction both in terms of *sentence constraints* on the final word but also of *semantic relatedness* between contextual information and the representation of the final word. These two factors have to be considered together in order to understand the amount of facilitation produced by the contextual sentences. Further evidence comes from the electrophysiological literature. The use of event-related brain potential (ERP) techniques shows the strong effect of contextual information. As discovered by Kutas and Hillyard (1980), the N400 component is highly sensitive to semantic information like word frequency, repetition, concreteness, semantic incongruities and contextual effects. This component indicates a negative peak in the ERP recording around 400 ms after the word onset. Moreover, Kutas and Hillyard (1984) identified a strong correlation between the amplitude of this component and the cloze probability. Federmeier (2007) describes the processes of the human brain that deal with language processing as "thinking ahead": the brain pre-activates some information associated with the most probable upcoming stimuli in order to be ready to process them.

Overall ERP studies have provided evidence for the facilitation effect as guided by context not only on very highly predictive words, but also on moderately predictive words (Jordan & Thomas, 2002; Kutas & Hillyard, 1984; Rayner & Well, 1996).

In order to better understand the nature of the information that is pre-activated by context, DeLong, Urbach, and Kutas (2005) used the N400 to analyse the indefinite articles before a more or less expected English word.

(7) The day was breezy so the boy went outside to fly ...

According to the cloze probability, the best completion for sentence (7) is "a kite". The alternative completion "an aeroplane" is still plausible but less probable. The authors claim that the expected higher N400 for the word "aeroplane" could be explained in terms of surprise (in the predictive account) but also in terms of the greater difficulty of integrating the word into the sentence representation (in the integration view). However, the integration problem should not affect the preceding article: the meaning and the complexity of the articles “a” and “an” are the same. The outcome of this experiment is in favour of the predictive account. They show that the same pattern related to the noun (higher N400 when the word is less expected) is also related to its article (higher N400 when the article does not match with the most expected word). This study confirms the results of previous experiments in favour of the predictive account and it also shows that the facilitation effect does affect upcoming words in general (and not only semantically rich words). Grammatical gender studies also reveal similar results (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004).

2.4 Semantic Feature Overlap

As suggested in Schwanenflugel and LaCount (1988) and Schwanenflugel and Shoben (1985), a strongly constraining context generates a stronger featural restriction process. This allows the activation of a very small and specific amount of meanings for the upcoming words. In this way, the facilitation affects only those words. On the other hand, weakly constraining contexts would allow facilitation to affect also less related/expected words. Moss (1997) performed three cross-modal priming experiments to study the access of word meanings over time in spoken word recognition. The author showed the activation of the mental representation of a word before it can be univocally recognised (before its “isolation point”). Moss showed that this activation was semantically driven and not simply due to associative priming. Moreover, the author identified some properties (functional properties: prototypical actions, actors, locations) that played an earlier role during the activation process. Other studies have shown a facilitatory effect exerted by unexpected words that are semantically related to the most expected word, even when those completions generate implausible sentences (Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984). As shown in Duffy et al. (1989) the semantic association between contextual words also produces greater facilitation.

With the related anomaly paradigm, Kutas and colleagues compared the N400 effect with congruent completions, incongruous completions, and incongruous completion semantically related to a congruent word (Kutas & Hillyard, 1984; Kutas et al., 1984). These experiments show that the overlap of some semantic information facilitates the processing of the upcoming words. Federmeier and Kutas (1999) investigated the relation between context and the structure of semantic memory during a reading task. The results were replicated also in a listening task (cf. Federmeier, McLennan, De Ochoa-Dewald, & Kutas, 2002). They analysed the effect of previous linguistic information in creating a specific expectation on the upcoming word. They also studied this expectation in terms of *semantic feature overlap* between the presented and the expected word. The authors recorded event-related brain potentials when participants were reading pairs of sentences such as:

(8) They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of ...

The first sentence creates the expectation about the upcoming information; the second one ends with the most expected word (e.g. "palms"), or an unexpected word from the same semantic category as the expected one (e.g. "pines"), or an unexpected word from an unexpected semantic category (e.g. "tulips").

Overall, the results show that the amplitude of N400 increases when the amount of expectation decreases: the expected word exerts a smaller N400 effect than the unexpected word from the same semantic category (within-category violation). The latter exerts a smaller N400 than the unexpected word from an unexpected semantic category (between-category violation).

These studies show that facilitation is driven by the strength of the prediction and the amount of semantic overlap between the expected and the realised word (Federmeier et al., 2007).

2.5 Conclusion

In this chapter we reviewed the most important studies on the effect of context on word processing. We reported the two main accounts that describe the effect of context on the activation of word meaning. The integrative account suggests a post-lexical influence of context while the predictive account indicates that context immediately pre-activates word meanings and this effect persists over time. We also showed that the facilitation of context is driven by frequency, sentence constraints and feature overlap.

The aim of this thesis is not to contrast the claims made by the integrative account and the predictive account. Based on the evidence provided by the ERP and the eye-tracking literature, we acknowledge the validity of the predictive account and we conduct a series of experiments meant to further investigate the incremental effect of context on word processing over time.

In Chapter 3, we discuss how the meaning of a word can be described in terms of semantic properties and how we can extract these properties to directly test the feature overlap theory. In Chapter 5 we study the effect of context over time described in the predictive account, by manipulating the amount of biasing information provided by the context. In Chapter 6 we address the feature overlap theory analysing the similarity between representations of word meanings generated by a distributional semantic model. Finally, in Chapter 7 we show the effect produced by semantic similarity relations on word processing.

Chapter 3

A Window into Word Meaning

In the previous chapter we discussed how contextual information exerts a facilitation effect on the access and processing of word meaning. We showed that different theoretical positions have been taken over time. According to the predictive account, the *feature overlap theory* is a plausible way of describing the context-word interaction: as the context unfolds, an increasing amount of semantic features become active; the words that match with those features are pre-activated and processed more quickly when encountered. In order to explain this process, in this chapter we discuss how words are represented in long term memory and how semantic similarity allows access to this information.

In Section 3.1 we show semantic feature norms can be used as a tool to describe the internal structure of word meaning. In Section 3.2 we present distributional semantics as a way to represent word meaning in terms of the contextual information those words occur with. Moreover, we discuss how the representations generated by distributional semantic models are cognitively plausible in different tasks. Traditional distributional models do not provide any information about the internal structure of word meaning. We show that there are some models that can collect this information. Finally, in Section 3.3, we report two experiments that test the cognitive validity of these models.

3.1 The Internal Structure of Word Meaning

For many years, researchers in different disciplines have investigated the mental representation of word meanings. In every day life, human beings are exposed to a huge amount of information about concrete and abstract entities. The way this information is acquired, stored and recovered from memory has been widely discussed (Murphy,

Property	Feature Type	Frequency
has_stripes	external_surface_property	22
a_carnivore	superordinate	17
an_animal	superordinate	15
has_teeth	internal_component	15
lives_in_jungles	location	14
a_feline	superordinate	13
is_black	external_surface_property	10
is_dangerous	systemic_property	10
lives_in_Africa	location	10
a_cat	coordinate	9

Table 3.1: The top 10 properties for the word "tiger" with the corresponding feature types and the production frequency (Source: McRae, Cree, et al., 2005).

2002). Nowadays, several theories of semantic representation describe word meaning as *featural primitives* (see Cree & McRae, 2003; McRae, Sa, & Seidenberg, 1997; Murphy, 2002; Vigliocco, Vinson, Lewis, & Garrett, 2004 for an extended discussion). These primitives are, for example, taxonomic relations and physical aspects of the entity considered that describe the meaning of a specific word. A well recognised tool to access and extract this information is the use of semantic feature norms (McRae et al., 1997; Rosch & Mervis, 1975; Vigliocco et al., 2004). To produce these norms, participants are asked to list the most salient attributes of a specific word. At the end of the collection, the attributes are normalised, classified according to a specific taxonomy, and numerically described (e.g. number the raw items, word frequency). For example, table 3.1 displays the 10 most frequent properties for the word "tiger" collected by McRae, Cree, et al. (2005). As shown in the example, semantic feature norms describe different aspects of word meaning: e.g. visual information (e.g. has stripes), taxonomic relations (e.g. a carnivore), and locations (e.g. lives in Africa).

Even though semantic feature norms do not constitute a precise record of semantic representation, they provide a window into different aspects of the internal representation of meanings (Medin, 1989). These properties have been used to describe various linguistic phenomena such as: the testing of semantic theories (Wu & Barsalou, 2009),

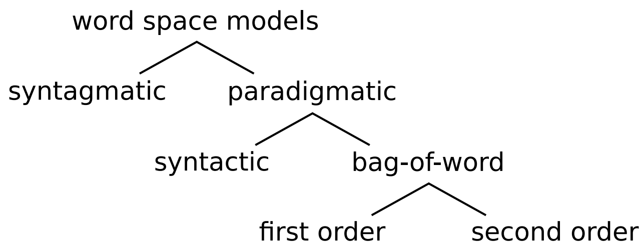


Figure 3.1: Taxonomy of different types of distributional semantic models (Source: Peirsman et al., 2008, p.908).

the computation of semantic similarity priming (McRae et al., 1997), and the performance of lexical decision tasks (Pexman, Lupker, & Hino, 2002). One of the major limitations of these collections is the extremely long process required to collect the norms for a relatively small set of meanings: e.g. the largest collection is the McRae, Cree, et al. (2005) one, which covers 541 living and non-living entities.

3.2 Distributional Semantics

Distributional semantics provides a complementary way to describe word meaning from a usage-based perspective. The meaning of a word can be defined in terms of the contexts in which the word occurs. This definition of word meaning refers back to the Wittgenstein's claim that "the meaning of a word is its use in the language" (1953, p.43). Distributional Semantic Models (DSMs) represent the meaning of a word in terms of the linguistic context the word is surrounded by. Consequently, two words are semantically similar if they occur in similar contexts. This assumption is known as the *Distributional Hypothesis* (Firth, 1957; Harris, 1954).

DSMs represent words as multidimensional vectors: each component of the vectors corresponds to another word (or a larger linguistic unit) which can be found in the contextual proximity of the target words (Sahlgren, 2008; Turney & Pantel, 2010).

DSMs can be classified according to the type of contextual information they extract and on the strategy they adopt to capture word meaning.

The taxonomy sketched in Figure 3.1 comes from Peirsman et al. (2008) and highlights the main differences between DSMs. As discussed in Sahlgren (2008), it is possible to classify DSMs based on their take on the extraction of contextual information. According to Sahlgren, DSMs are either syntagmatic or paradigmatic. This distinction is taken directly from de Saussure's definition of paradigmatic and syntagmatic re-

	Zoo	Mammal	Road	Africa	Ride	Hunt	Run	Tail
Tiger	1	1	0	0	0	1	1	1
Crocodile	1	0	0	1	0	1	1	1
Bicycle	0	0	1	0	1	0	1	0

Table 3.2: Toy vectors representing the target words: “tiger”, “crocodile” and “bicycle”. Each dimension contains binary values that indicate the co-occurrence of the words in the column headers with the target words.

lations between co-occurring words (Saussure, 1915). Syntagmatic relations hold between co-occurring words, while paradigmatic relations hold between words that never co-occur but share the same linguistic context. Distributional models can be classified in the same way. Syntagmatic models are those that compute the frequency of a word occurring in a certain region of text (often a document). Meanings are represented in a words-by-documents co-occurrence matrix. In the representation encoded in syntagmatic DSMs, two words are similar if they occur in the same documents. The most famous example of a syntagmatic model is the Latent Semantic Analysis by Landauer and Dumais (1997). On the other hand, paradigmatic models consider two words similar when they share the same collocates. In the case of paradigmatic DSMs, meanings are represented in a words-by-words co-occurrence matrix. An example of a paradigmatic model is the Hyperspace Analogue to Language algorithm by Lund and Burgess (1996). Paradigmatic models can be further classified into syntactic and bag-of-words models. Syntactic models use linguistic knowledge to produce more specific representations of word meaning. For example, Lin (1997) and Padó and Lapata (2007) parsed the corpus to extract context words syntactically related to the target word. Bag-of-words models collect all the context words within a certain window without any prior knowledge about these words. Bag-of-words models can be further classified into first-order and second-order models. First-order models are those that directly record the context around the target word (e.g. Mitchell & Lapata, 2008). Second-order models record the context words that surround the first-order context words (e.g. Schütze, 1998). The model we use in Chapter 6 is a first-order bag-of-word model developed by Mitchell (2011). To clarify how bag-of-words DSMs work we constructed a toy example reported in Table 3.2.

Table 3.2 displays the toy vectors representing the target words: “tiger”, “crocodile”, and “bicycle”. In this case, each vector dimension contains a binary value (0 or 1). This value is updated to 1 if the words in the column headers appear in the contextual proximity of the targets. The vectors representing “tiger” and “crocodile” share more dimensions (zoo, hunt, run, tail) than the vectors representing “tiger” and “bicycle” (run). Consequently, the semantic similarity between “tiger” and “crocodile” is higher (0.8) than the semantic similarity between“tiger” and “bicycle” (0.26).

3.2.1 The Cognitive Validity of Distributional Models

Over the years, DSMs have been used to describe a large amount of different linguistic and psychological phenomena. For this reason, Lenci (2008) highlights the importance of distinguishing between two versions of the Distributional Hypothesis: a strong and a weak version. The main difference between the two versions is related to the cognitive validity assigned to the representations generated.

The “weak version” of the Distributional Hypothesis recognises only the quantitative use of DSMs to describe semantic relations. On the other hand, the “strong version” posits the cognitive validity of those models: DSMs provide an insight into the internal representation and structure of the lexicon in the brain. In recent decades, DSMs have become very popular in psycholinguistics where they are used to successfully model different aspects of human language acquisition and processing such as: vocabulary acquisition (Landauer & Dumais, 1997), semantic similarity judgements (McDonald, 2000), category-related deficits (Vigliocco et al., 2004), and semantic priming (Lapesa & Evert, 2013; Padó & Lapata, 2007). Moreover, in some studies DSMs are used as a complementary tool to experimental data (Andrews, Vigliocco, & Vinson, 2009).

3.2.2 Extracting Feature Norm-like Data from Corpora

The information extracted from corpora and used to construct word vectors is very rich and provides a powerful instrument to compare the meaning of different words. However, it does not provide any information about the internal structure of word meanings in terms of *interpretable dimensions* (Murphy, 2002, p.430). The main question is: are the words captured by the model as vector dimensions comparable to the semantic properties produced by humans when describing a specific meaning?

In the last decade, researchers have developed systems that aim to extract feature norm-like data from corpora (Almuhareb & Poesio, 2005; Baroni & Lenci, 2010; Baroni et al., 2010; Devereux, Pilkington, Poibeau, & Korhonen, 2010).

In this section we describe Strudel (STRUctured Dimension Extraction and Labelling), a distributional semantic model representing word meaning in terms of *weighted interpretable typed properties* extracted from corpora (Baroni et al., 2010). We use the output of this model to construct the linguistic context for the visual world experiment reported in Chapter 7. The difference between Strudel and the traditional DSMs is that it analyses the semantic relations between the possible context words and the target. These relations are described in terms of linguistic patterns that connect two words. Only the context words following certain predefined patterns are recognised as good candidates and become part of the target word representation. The authors claim that these words can be considered semantic properties of the target word. The properties are nouns, adjectives and verbs that occur in a pre-defined window surrounding the target word. The model also stores the links connecting the properties with the corresponding target. The amount of different links connecting two words is the criterion according to which the rank of the combination of a property and the target word is computed. The assumption behind this approach is that a higher amount of links connecting two words indicates a stronger semantic relation between them. To clarify the way Strudel describes word meaning, in Table 3.3 we report the six most frequent properties produced by Strudel for the word "tiger" and the relations connecting the target word (C) with its property (P) (e.g. "tiger_in_jungle"). The properties are ranked based on the amount of links connecting them to the target (as indicated by the percentages in brackets). As shown in this example, the model can successfully identify locations (e.g. "jungle" and "zoo"), semantic relations ("lion"), actions ("maul" and "kill") and also visual features ("stripe"). Baroni et al. (2010) show that Strudel outperforms traditional DSMs on different tasks targeting a meaning representation in terms of semantic features. Strudel successfully identifies salient properties that correlate with human generated feature norms (from McRae, Cree, et al., 2005). It also groups these words by property type and produces superordinate clusters of words based on the property distribution.

3.3 Semantic Relations and Eye-Movements

One of the common principles shared by feature norms and distributional semantics is that the similarity between meanings is reflected in the amount of shared features. Cooper (1974) found that participants, when hearing the word "Africa", are more likely

Property	Type Information
jungle-n	C_in_P (53%), C_through_P (11%)
zoo-n	C_in_P (60%), C_from_P (10%)
lion-n	P_on_C (46%), C_because_P (15%)
maul-v	P_by_C (47%), C_P (47%)
kill-v	C_P (51%), P_C (25%), P_by_C (19%)
stripe-n	P_of_C (23%), P_on_C (23%) C_have_P (15%), C_with_P (15%)

Table 3.3: The top 6 properties produced by Strudel for the word "tiger" and the link (typed information) between the word (C) and the property (P) (Source: Baroni et al., 2010, p.236).

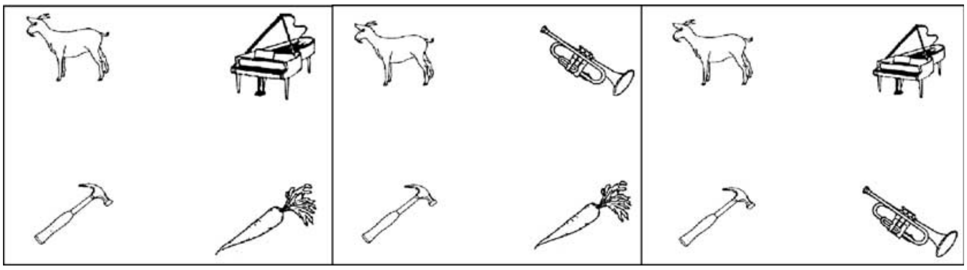


Figure 3.2: Three scenes (from left to right: target, competitor, and target & competitor condition) for the word "piano" (Source: Huettig & Altmann, 2005, p.B25).

to fixate semantically related objects such as the picture of a lion, a zebra or a snake (cf. Section 4.2). Huettig and Altmann (2005) show that eye-movements are driven by the semantic similarity between a word and the mental representation of an object depicted on the screen. Figure 3.2 displays an example of the three scenes associated to the linguistic stimulus "piano". For each linguistic stimulus, the authors also identified a semantically, but not associatively, related competitor (here, "trumpet").

The first scene on the left (the target condition) depicts a piano and three distractors. The central scene (the competitor condition) shows a trumpet and three distractors. Finally, the right scene (the target & competitor condition) shows both a piano and a trumpet and two distractors. The authors analysed the eye-movements after the onset of the critical word (here, "piano") in the three conditions. The results show that in the first two conditions when hearing the word piano there is an increasing amount of fixations on the image of the piano or of the trumpet. When, as in condition three (target

& competitor), both target and competitor are depicted, the higher amount of fixations is on the piano. However, the fixations on the trumpet are higher than the fixations towards the two distractors. Moreover, Huettig and Altmann (2005) show a correlation between the probability of fixating the competitor and the similarity measure based on semantic feature norms (Cree & McRae, 2003). The authors claim that the degree of semantic overlap between the linguistic stimulus and the mental representation of the objects depicted on the screen affects participants' eye-movements.

Huettig et al. (2006) wanted to provide evidence for the cognitive plausibility of distributional semantic models in capturing semantic similarity. The visual and linguistic stimuli developed for this experiment are used in Experiment 2 (only visual stimuli), Experiment 5 (visual and linguistic stimuli), and Experiment 6 (visual and linguistic stimuli), as described in the next chapters of this thesis. For this reason, we provide a more detailed description of the stimuli and the design of this experiment.

Twenty-six target-competitor pairs of words were used for this experiment. The words in each pair were semantically but not associatively related. In order to minimise the associative relations, the authors consulted two British word association norms (Kiss, Armstrong, Milroy, & Piper, 1973; Moss & Older, 1996). The semantic relation among the words in each pair was computed using two distributional models: the LSA model (Landauer & Dumais, 1997), and the contextual similarity model (McDonald, 2000).

Traditionally, association norms are collected asking participants to produce the first word which first comes to their mind when hearing another word (Kiss et al., 1973). While, semantic norms provide a richer description of word meanings (e.g. semantic feature norms). The difference between association and semantic relations has been widely discussed in the literature (Hutchison, 2003). For example, Yee, Overton, and Thompson-Schill (2009) show that in a visual world experiment eye-movements are driven by semantic relations (ham-eggs) but not by simple associations (e.g. iceberg-lettuce). Similarly, McRae and Boisvert (1998) show that automatic semantic priming occurs also when there is not association between the prime and the target.

In the experiment performed by Huettig et al. (2006), each word was embedded in a neutral sentence. For example, for the target/competitor pair of words toaster /corkscrew the linguistic stimuli were:

(1) `Target`: First, the man disagreed somewhat, but then he noticed the **toaster**

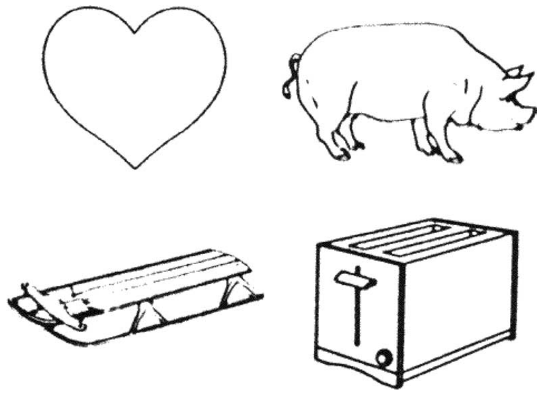


Figure 3.3: Example scene for the pair toaster (target)/corkscrew (competitor) (Source: Huettig et al., 2006, p.71).

and appreciated that it was useful.

(2) `Competitor`: First, the man disagreed somewhat, but then he noticed the **corkscrew** and appreciated that it was useful.

As shown in Figure 3.3, the visual scene contains the representation of the target object (here, "toaster") and three distractors. In this experiment the competitor object (here, "corkscrew") was never depicted. The visual scene was an array of black and white line drawings taken from the Snodgrass and Vanderwart (1980) collection. A series of norming studies were performed on the visual objects in order to reduce any effect not directly related to semantic similarity: picture naming agreement, image agreement, familiarity, visual complexity, and word frequency for the name. In order to exclude phonological competition, the names of the target and distractors objects were starting with different phonemes. Moreover, the authors performed a study on the visual similarity between the target object and the mental image of the competitor object. In this way they could isolate the effect of visual similarity from the semantic relation between target and competitor.

Figure 3.4 shows the probability of fixating the target object on the screen when listening to the target word or to a competitor word. All the fixations are aligned at the onset of the critical word. The plot shows that when listening to a target word (e.g. "toaster") the amount of fixations towards the target object increases over time. Similarly, when listening to a competitor word (e.g. "corkscrew") subjects were still looking more often towards the target object than the other three distractors. The authors analysed the correlation between the probability of fixating the target word in

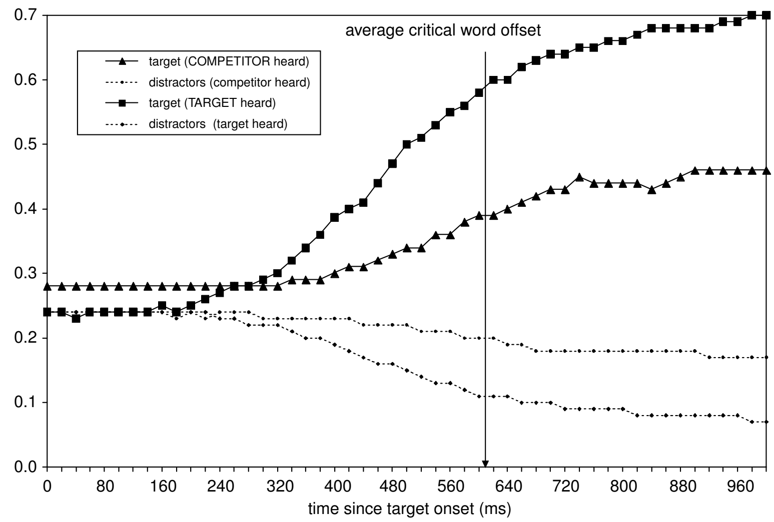


Figure 3.4: Fixations over time for the target and competitor conditions (Source: Huettig et al., 2006, p.73).

the competitor condition with the LSA scores ($r = 0.42, p = 0.033$) and the contextual similarity scores ($r = 0.58, p = 0.002$). In both cases a moderate correlation emerged.

Two main conclusions were drawn from this experiment. First of all, they replicated the outcomes of Huettig and Altmann (2005) showing that semantic similarity plays an essential role in accessing word meaning: the facilitation effect is not only driven by associative priming (as suggested by Hutchison, 2003) but also by semantic similarity by itself. Moreover, they computed the correlation between eye-fixations and semantic similarity proving the psychological validity of those models in describing semantic relations. They found no difference between the two models.

3.4 Conclusion

In this chapter we described word meaning in terms of semantic properties. In Section 3.1, we showed that semantic feature norms have been widely used to describe different linguistic phenomena. In Section 3.2 we presented DSMs as a complementary way of representing word meaning. One of the limitations of DSMs is that they do not provide a representation of the internal structure of meanings. A possible solution is the use of Strudel, a distributional semantic model that represents word meaning in terms of weighted interpretable typed properties. Finally in Section 3.3 we reported two experiments meant to support the cognitive validity of describing semantic similarity in terms of feature overlapping and the use of distributional semantic models to compute those similarities.

In the next chapters we extend the outcomes of the experiments described here in different ways. In Chapter 6 we test the ability of a bag-of-words distributional semantic model to correctly describe the semantic relation between an increasing amount of biasing contextual information and the target word. We evaluate the cognitive validity of these representations by comparing the results with reading data. In Chapter 5 we use the visual scenes of Huettig et al. (2006) (see section 3.3) to analyse the effect of an incremental amount of biasing information on accessing word meaning. The use of the same normed stimuli as Huettig et al. (2006) will reduce possible effects not directly related to semantic similarity. The same stimuli are also used to design a new visual world experiment described in Chapter 7. This study directly replicates the original experiment, adding an extra variable to the analysis: the semantic relation between the context and the target word. We constructed the linguistic contexts using properties generated by Strudel that are associated to the target word (either the target or the competitor condition). This manipulation allows both to test the ability of a distributional model to provide plausible information for building linguistic contexts, and to compare the effect of a biasing context on semantic similar words.

Chapter 4

Methodology

In this chapter we describe the use of reading and eye-tracking experiments to study language processing over time. In the experiments reported in Chapter 5 and in Chapter 7, we use the self-paced reading paradigm to analyse the effect of context on the reading time of the target word at the end of the sentence. Moreover, we use the visual world and the blank screen paradigms to analyse the predictive effect of the unfolding contextual information. Specifically, Experiment 1 (Section 5.2) and Experiment 3 (Section 5.4) are self-paced reading studies, Experiment 2 (Section 5.3) and Experiment 5 (Section 7.1) are visual world paradigm studies, while Experiment 6 (Section 7.2) is a blank screen paradigm study.

4.1 The Self-Paced Reading Paradigm

Self-paced reading (SPR) is a well established method to study language processing on-line. This methodology records the time required by the participants to read at their own pace a certain segment of the sentence (words, multiple words, or phrase). The reading times (RTs) reflect the time required by the participants to process the linguistic information provided (Just, Carpenter, & Woolley, 1982). This technique tries to make the experience as similar as possible to normal reading.

For the first time, Just et al. (1982) used this paradigm to analyse the connection between reading times and language processing at sentence level. Jegerski and VanPatten (2013) report different studies where the SPR paradigm has been used: the analysis of the time required to process ambiguities (e.g. the garden path phenomena in Trueswell & Kim, 1998), anomalies (e.g. number agreement in Foote, 2010) and distance dependencies (e.g. broken agreement in Jiang, 2004).

Different designs fall under the same definition of SPR. Overall, it is possible to group various experiments based on two main differences. The first difference is between linear and centred designs. In the former, the entire sentence is always displayed (dashed) on the screen. In the latter, only one word (or segment) at a time appears on the centre of the screen. The second difference is between cumulative and non-cumulative designs. The former leaves uncovered the part of the sentence that has already been processed. The latter dashes again the processed information leaving un-dashed only one word at a time. The choice of the dimension of the segment showed and the design adopted are key factors to consider when preparing a SPR experiment. Overall, the linear non-cumulative design (also called moving window design) is considered as the most natural one because participants have a clear perception of the evolution of the sentence and the reading proceeds naturally from left to right. At the same time, participants cannot develop any reading strategy (Ferreira & Henderson, 1990).

Now we describe the design of a linear non-cumulative SPR experiment. At the beginning of each trial, the entire sentence is on the screen but fully dashed. The experiment starts when the participant presses a button. Every time the button is pressed participants can read the next segment of the sentence while the previous one is dashed again. Participants can read only one segment at a time. The time occurring between two button presses is recorded as the reading time for the given segment. Figure 4.1 shows an example of this design.

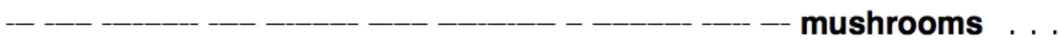


Figure 4.1: Example of what participants see on the screen during a moving-window self-paced reading study. The complete sentence taken from the stimuli of Experiment 1 (see Section 5.2) is: “In the forest the picker was holding a basket full of mushrooms carefully”. The dots indicate the continuation of the sentence and do not appear during the experiment.

Even though the advantages of using this methodology are clear there are also some downsides to consider. Performing a SPR experiment is relatively cheap and it does not require any specific equipment (except for a computer to run it) nor the continuous supervision of the experimenter (Jegerski & VanPatten, 2013). However, presenting one word after the other while pressing a button is relatively unnatural. Rayner and Clifton (2002) shows that the RTs in a SPR experiment are longer than normal reading times.

4.2 Eye-Movements and Language Processing

In the previous section we discussed some downsides in the use of the SPR paradigm: in particular the fact that the procedure is not completely natural. A possible solution to overcome the limitation is the use of eye-tracking techniques (Holmqvist et al., 2011). The advantage of this measures is that participants can be involved in more natural tasks (e.g. listening to a sentence) while their eye-movements are recorded. Eye-movements are spontaneous and automatic and do not require any voluntary action performed by the subject (like pressing a button in the SPR studies).

During the analysis of the recorded data, two main types of eye-movements are usually considered: saccades and fixations. Saccades are rapid movements of the eyes (around 30ms to 80ms) between two points in the visual space. Researchers normally consider that humans are almost blind during these movements. Fixations occur between saccades when the eyes are relatively stable. During fixations micro-movements (as microsaccades, tremors, and drifts) take place. The position of the eyes during fixations shows the visual region where the attention is allocated. While the duration of the fixations is an indicator of the cognitive load and fluctuates between 200ms and 300ms according to the task. For example, in a reading task, a fixation lasts around 250ms on average: the position of the eyes indicates the word that is read in a specific time, while the fixation duration indicates the occurring lexical processing (Sereno, 2003).

In the next sections, we report two experimental paradigms that use eye-movements to study the processing of linguistic information over time: the visual world paradigm and the blank screen paradigm.

4.2.1 The Visual World Paradigm

Cooper (1974) was the first to describe the strong relation between language processing and visual attention. The author showed that, when listening to a sentence, listeners' eye-movements are closely synchronised to the linguistic stream. Specifically, participants were more likely to look at the picture of a lion than at the other pictures on the screen (the distractors) when they were hearing the word "lion". Cooper also showed that participants looked more often to the picture of a lion, a zebra or a snake when hearing the associated word "Africa". With this study, the author not only evidenced a strict relation between eye-movements and language processing but also that eye-movements are guided by more subtle linguistic relations (e.g. semantic associations).

Only twenty years later with the work of Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995), the visual world paradigm became widely recognised for the study of the integration of linguistic and visual information. Tanenhaus et al. (1995, p.1632) showed that "eye movements provide insight into the mental processes that accompany language comprehension". They demonstrated that eye-movements are time-locked to the linguistic information provided. When participants hear the word that allows the identification of the target object on the screen, their eyes automatically move to the identified target object. Similarly, the authors showed that visual information is combined with linguistic information to solve linguistic ambiguities.

The visual world paradigm was originally associated to language comprehension experiments, however is has been widely used also to study language production. In production studies, participants are exposed only to the visual stimulus while they have to perform a production task. A traditional task consists in the description of the scene depicted on the screen. Eye-movements reveal the order in which the visual information is processed and how this information affects the production of the linguistic description (e.g. Griffin & Bock, 2000).

In a traditional language comprehension setup, participants listen to a pre-recorded linguistic stimulus when they are looking to a visual scene on a computer screen. Their eye-movements are recorded and time locked. In this way it is possible to study the correspondence between the unfolding linguistic stimulus and the fixations towards specific objects depicted on the screen. Some of these objects are directly related to the linguistic stimulus (target objects) while others are distractors. The visual scene can be composed by arrays of black and white line-drawings (e.g. Huettig et al., 2006), semi-realistic scenes (e.g. Altmann & Kamide, 1999), realistic scenes (e.g. Staub et al., 2012), or printed words (e.g. Huettig & McQueen, 2007). Figures 4.2, 4.3, 4.4, and 4.5 show an example for each type of visual scene. The choice of a specific type of scene is based on the research question the experimenter wants to address. For example, as discussed in Huettig, Rommers, and Meyer (2011), the use of realistic or semi-realistic scenes makes possible the study of how participants perceive and process the relation between different objects in the scene. While the use of arrays of objects addresses more specifically the activation and processing of word meaning driven by the linguistic information.

In the analysis of visual world data, the dependent variable is often the probability

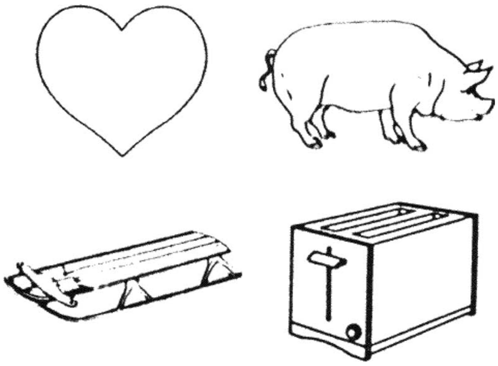


Figure 4.2: Example of a scene composed by an array of black and white line-drawings (Source: Huettig et al., 2006, p.71).

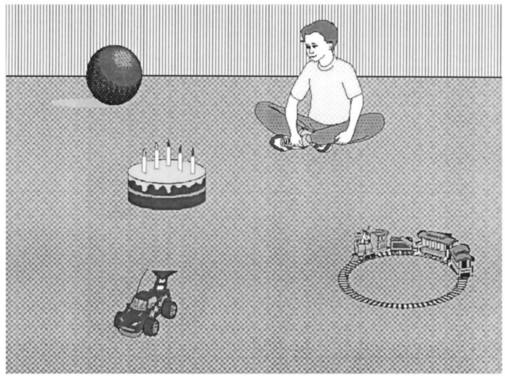


Figure 4.3: Example of a semi-realistic scene (Source: Altmann & Kamide, 1999, p.250).

Figure 4.4: Example of a realistic scene (Source: Staub et al., 2012, p.924).

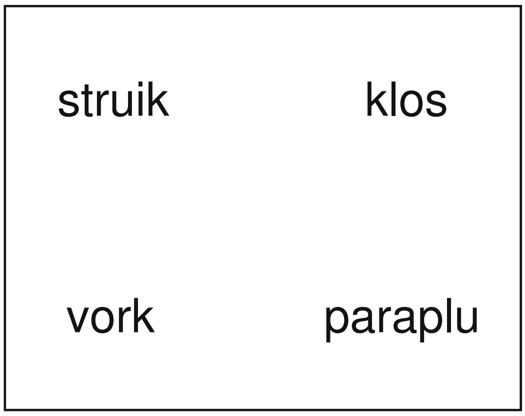


Figure 4.5: Example of a scene composed by printed words (Source: Huettig & McQueen, 2007, p.473).

of looking at a specific area of interest (AOI) in the visual scene over time. The AOI in the visual scene corresponds to one or more visual elements related to the linguistic input.

In order to clarify the traditional procedure required to perform a visual world study, we report the seminal study of Altmann and Kamide (1999). In their analysis, the authors show that verbs restrict the fixations towards the possible post-verbal objects. Figure 4.3 shows an example of the scenes used in the experiment. Each visual scene is associated to two different sentences. For example, the sentences for Figure 4.3 are:

(1) the boy will eat the cake

(2) the boy will move the cake

The verb "eat" in sentence (1) imposes a selectional restriction to only one object in the visual scene. In this case only "cake" can fill the post-verbal object slot in the sentence. While the verb "move" in sentence (2) is compatible with all the objects depicted.

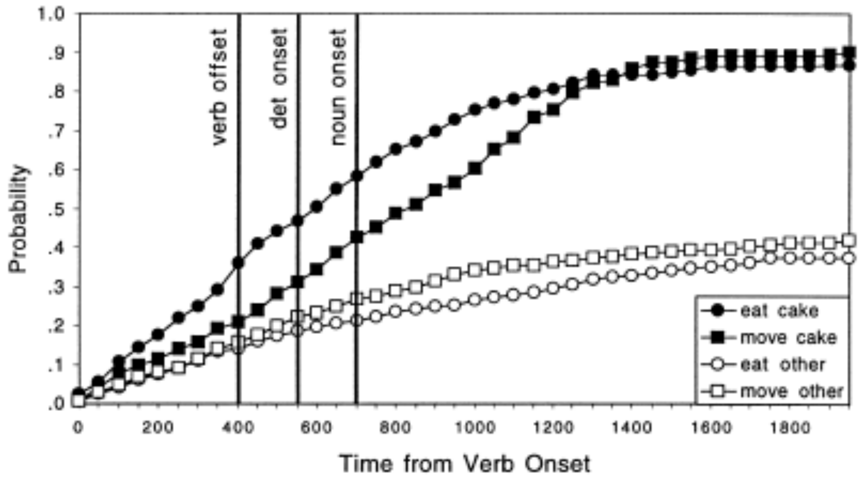


Figure 4.6: The cumulative probability of fixating the target object (cake) or the distractor object (other) as a function of the verb used ("eat", "move") (Source: Altmann & Kamide, 1999, p.253).

In order to define the AOI for the analysis, the verb onset and offset, and the post-verbal noun onset and offset have been marked. The plot in Figure 4.6 shows the cumulative probability of fixating the target object (cake) or one of the distractors (a toy train set, a toy car, and a balloon) after hearing the verb "eat" or the verb "move". The results indicate that the amount of fixations towards the target object are higher when the selectional restrictions imposed by the verb apply to only one possible upcoming object (the cake) than when they apply to more than one objects.

Overall, the results demonstrate that the selectional restrictions imposed by the verb guide the eye-movements towards the objects that respect these restrictions. The authors claim that eye-movements are mediated by the mental representation of the situation described in the linguistic context. The outcome of this experiment supports the predictive account described in Section 2.2: language processing is incremental, and the meaning of the upcoming word is pre-activated even before hearing the corresponding word.

4.2.2 The Blank Screen Paradigm

Altmann (2004) reports a study similar to the Altmann and Kamide (1999) one where the visual and the linguistic information do not co-occur simultaneously. When the experiment starts, participants see the visual scene on the screen. After 5 seconds the image disappears from the screen (blank screen) and the pre-recorded linguistic stimulus is played. Eye-movements are recorded in the second phase. The results obtained correspond with those in Altmann and Kamide (1999): at the time of the verb, participants look at the area of the blank screen in which the most expected post-verbal object was depicted. These results support the theory that the anticipatory eye-movements described in Altmann and Kamide (1999) are "dependent on a mental record of the scene that is independent of whether the visual scene is still present" (Altmann, 2004, p.B79).

4.3 Conclusion

In this chapter we described the methodologies used in the experiments reported in the next chapters of this thesis. In Section 4.1 we described the self-paced reading paradigm (Experiment 1 and Experiment 3). This paradigm is used to study the effect of context on the time required to process the target word appearing at the end of the sentence. In Section 4.2 we reported two techniques to study eye movements. The visual world paradigm (Experiment 2 and Experiment 5) makes possible the analysis of the effect of the unfolding context on the pre-activation of the upcoming target word. Finally, with the blank screen paradigm (Experiment 6) we replicate Experiment 5 separating the processing cost of the visual stimuli from the cost of the linguistic stimuli.

Chapter 5

Incrementality and Contextual Effects

5.1 Introduction

The aim of the experiments reported in this chapter is to study contextual constraints on word processing, focusing in particular on how context activates word meaning. According to Federmeier and Kutas (1999), contextual facilitation effects can occur in different tasks (and affect, e.g., reading times, lexical decision times, pronunciation times); effects appear both at the level of lexical priming and at the level of the entire sentence. Based on this assumption, we want to clarify the relation between single contextual words and the entire sentence. We constructed sentence materials that provide a differential number of context words that bias comprehension towards the target word. This design allows us to determine how such facilitation (or bias) occurs. The predictive account described in Section 2.2 suggests that context immediately affects the pre-activation of the meaning of the upcoming words, and that this effect is cumulative. In order to test this assumption, we propose two possible hypotheses. The *incremental activation hypothesis*, which directly reflects the predictive account, suggests that the degree of facilitation in word processing starts immediately and increases with the amount of context available. The *immediate activation hypothesis* states that a certain amount of biasing information is required to boost the activation of the upcoming words. Moreover, it suggests that once a sufficient amount of contextual support is reached, no more facilitation occurs. In Figure 5.1 and Figure 5.2 we sketch the two hypotheses: the two lines describe the facilitation effect while increasing the amount of context available. The two activation hypotheses described in this thesis are not extracted from the literature. As described in Chapter 2, many studies have analysed the effect of non biasing vs. biasing context on word processing. However, the authors of those studies never manipulated directly the amount biasing information in the linguistic context. This is the reason why we formulated the two hypotheses in order to describe the effect of various amount of context presented over time.

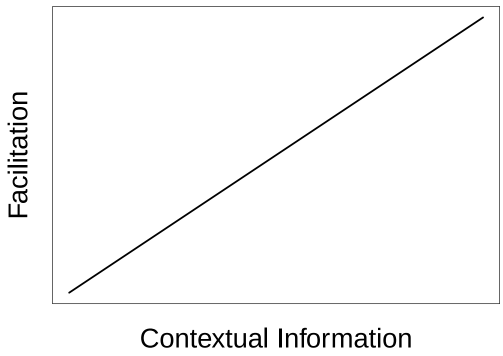


Figure 5.1: Schematic representation: the incremental activation hypothesis.

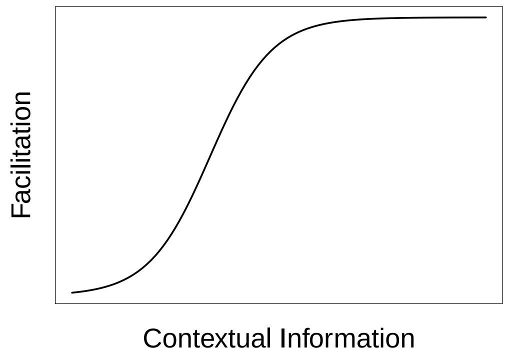


Figure 5.2: Schematic representation: the immediate activation hypothesis.

We performed four experiments to test these hypotheses. In a self-paced reading experiment (Experiment 1), we study if the amount of context available has an effect on the ease of word processing, measured as the reading time for the target word. In a visual world study (Experiment 2), we present the target word pictorially and we measure the degree of facilitation (i.e., the number of looks) that occurs while the context words are processed. The use of this paradigm gives us access to the time course of word integration. In another self-paced reading experiment (Experiment 3), we test the effect exerted by the order of contextual information on word processing. Finally, in a series of word association studies (Experiment 4) we analyse more directly the relation between different contextual words.

Experiments 1 and 2 appear in Frassinelli et al. (2013).

5.2 Experiment 1: Contextual Incrementality and Word Activation

The aim of this self-paced reading experiment is to analyse the effect that context information has on participants' reading time (RT) for the target word. As widely discussed in the reading literature (Morris, 1994), higher coherence between the context and the target word is reflected in lower RTs. We therefore predict that RTs for the target word are reduced in proportion to the number of context words presented because context words bias the reader towards the target word. The incremental activation hypothesis predicts that when we add a new HB word to the context, the amount of time required to read the target word decreases over time. The immediate activation hypothesis supports the idea that when the pre-activation of the target word has taken place, the inclusion of new HB words to the context does not significantly affect the RT of the target word.

5.2.1 Method

Materials We used the same 24 words as Huettig et al. (2006) but we dropped the semantically related condition, as it is not relevant for the present experiment (see Section 3.3). We embedded these target words into a sentential context using the following general structure:

(1) *location* – *actor* – verb – *object* – **target word** – spill-over region

The target word is in **bold**, the three context words in *italics* (see below for an example). For each target word, we identified three context words which were highly related to it (high-biasing (HB) words) and three context words that were semantically coherent with the target word but not strongly associated to it (low-biasing (LB) words). The following examples describe the resulting eight possible combinations of LB and HB context words:

(2) `All LB context` (*None*): On the *path*, the *man* was holding a *box* full of **mushrooms** carefully.

(3) `HB location context` (*Loc*): In the *forest*, the *man* was holding a *box* full of **mushrooms** carefully.

(4) `HB actor context` (*Act*): On the *path*, the *picker* was holding a *box* full of **mushrooms** carefully.

(5) `HB object context` (*Obj*): On the *path*, the *man* was holding a *basket* full of **mushrooms** carefully.

(6) `HB location and actor context` (*LocAct*): In the *forest*, the *picker* was holding a *box* full of **mushrooms** carefully.

(7) `HB location and object context` (*LocObj*): In the *forest*, the *man* was

holding a *basket* full of **mushrooms** carefully.

(8) `HB actor and object context` (*ActObj*): On the *path*, the *picker* was holding a *basket* full of **mushrooms** carefully.

(9) `All HB context` (*All*): In the *forest*, the *picker* was holding a *basket* full of **mushrooms** carefully.

This resulted in 192 experimental sentences: eight contexts for each of the 24 target words. Example (2) shows the condition with three LB properties. Examples (3)−(5) show the condition where only one contextual word is HB; while Examples (6)−(8) show the conditions with two biasing HB words in the context. Finally, Example (9) shows the condition with three HB words. The complete list of the linguistic stimuli for this experiment is reported in Appendix A.

Norming Studies In order to control the biasing effect of each context word, we conducted a series of norming studies on Amazon Mechanical Turk[1]. The participants were all native English speakers. Each participant was allowed to perform only one study.

Sentence Plausibility Study Forty participants assessed how plausible the experimental sentences were by rating them on a scale from 1 (completely implausible) to 7 (completely plausible). A sentence was considered plausible when the averaged rating was higher than 4. This process allowed us to identify those sentences that were not completely plausible; they were replaced and re-tested.

Sentence Completion Study Forty new participants evaluated the predictability of the target word from the sentence context. They had to complete each sentence (with the target word removed) by typing in a noun. Each sentence was completed by 5 subjects; for this reason the number of completions matching the target for each item goes from zero up to five. Table 5.1 reports the amount of answers for the expected completion averaged by the type of HB words that appear in the sentence and ordered by score. We included only the answers that perfectly matched with the removed target word. The increasing amount of HB information produces a higher amount of expected target words (expected answer). Only the condition *LocAct* shows a lower score than the conditions with only one HB property; this effect disappears in the next tasks.

[1]https://www.mturk.com

Table 5.2 lists the coefficients for the analysis performed with a linear mixed effects model (LME, version 0.9-2). Unless defined otherwise, all the analyses performed in this thesis report the results for the maximal model that includes the random intercept and slope for all the effects (Barr, Levy, Scheepers, & Tily, 2013). The analyses are performed with the software for statistical computing R[2] (version 3.0.1). The number of expected answers was the dependent variable, the different contextual factors were contrast coded against the reference level *None*; `Participant` and `Item` were included as random intercepts and slopes. The model shows that the difference between *None* and all the other levels is statistically significant: when performing a sentence completion study, participants are facilitate in identifying the expected target word when at least one context word is HB. From a purely descriptive perspective, the presence of a HB object in the context produces a higher amount of expected answers. There are three possible explanations for this result. The semantic category of the object is similar to the category of the target word (often an object too). The object and the target word are syntactically connected (e.g. a basket full of mushrooms). Moreover, the object is the context word that is closer to the target one. In a Tukey HSD post-hoc analysis we compared the differences among levels that share the same amount of HB words. The difference between the conditions with one HB word does not reach significance ($p > .05$); while the condition *LocAct* is significantly different both from *ActObj* ($p < .05$) and from *LocObj* ($p < .05$). This difference is due to the very low score reported for *LocAct*.

For a more general discussion, we averaged the amount of expected answers from Table 5.1 by the number of HB words. Figure 5.3 shows the average number of expected answers grouped by number of HB words (zero, one, two, and three HB words). Descriptively, the number of expected answers increases with the amount of HB words provided: 0.32 ($\pm .17$) out of 5 after three LB words (condition *Zero*); 1.15 ($\pm .18$) after one HB word (condition *One*); 1.51 ($\pm .19$) after two HB words (condition *Two*); 2.16 ($\pm .34$) after three HB words (condition *Three*). The LME analysis is reported in Table 5.3. The number of expected answers was the dependent variable, the number of HB words was the factor of the analysis and was contrast coded against the reference level *Zero*; `Participant` and `Item` were included as random intercepts and slopes. The number of expected answers driven by three LB context words is statistically lower than the number of expected answers when the context contains one or more HB words. In a post-hoc analysis we compared pairwise the differences between

[2]http://www.r-project.org

Condition	Score
None	0.32 ± 0.17
Loc	0.96 ± 0.34
Act	1.00 ± 0.28
Obj	1.48 ± 0.33
LocAct	0.72 ± 0.21
ActObj	1.88 ± 0.36
LocObj	1.92 ± 0.35
All	2.16 ± 0.34

Table 5.1: Sentence completion study: Number of completions matching the target words (out of 5) with standard errors. The results are grouped by number of HB words and ordered by score.

Predictor	Coefficient
(Intercept)	0.32
Loc	0.64^{*}
Act	0.68^{*}
Obj	1.16^{***}
LocAct	0.40^{*}
ActObj	1.56^{***}
LocObj	1.60^{***}
All	1.84^{***}

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.2: Sentence completion study: LME coefficients for the scores in Table 5.1.

the four conditions. The analysis shows that the difference between condition *One* and *Three* is significant ($p = .032$). While the difference between the remaining conditions (*One-Two* and *Two-Three*) does not reach significance ($p > .05$). The analysis indicates that when at least one HB property is in the context, the inclusion of one more HB word does not significantly facilitate the identification of the target word.

The sentence completion study reported in this section was conducted primarily to test the quality of the linguistic stimuli we used for the self-paced reading experiment. Nonetheless, it provides some interesting evidence on the effect of context words in facilitating the identification of the expected target word. The analysis shows that the presence of at least one HB word in the context significantly facilitates the production of the expected word. Moreover, the inclusion of an additional HB word after a HB word does not generate significant differences in the amount of expected answers produced.

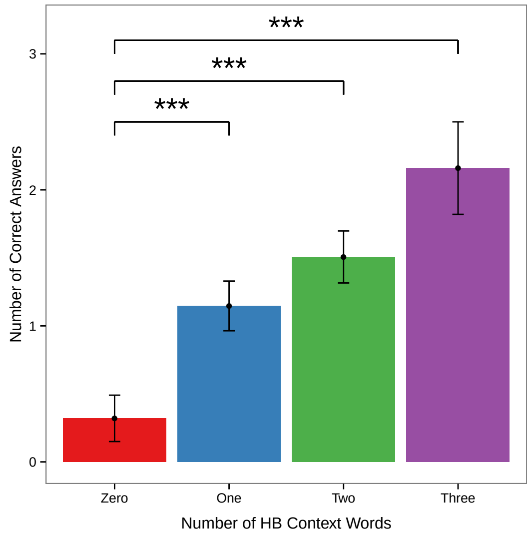


Figure 5.3: Sentence completion study: Plot of the number of expected answers (out of 5) averaged by the number of HB context words.

Predictor	Coefficient
(Intercept)	0.320
One	0.827***
Two	1.187***
Three	1.840***

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.3: Sentence completion study: LME coefficients for the scores in Figure 5.3.

Word Completion Study Forty new participants performed a word completion task. Three context words appeared on the screen one after the other. Participants had to type the word more related to these context words. The aim of this study was to exclude possible syntactic effects (word order, but also the effect of the verb) occurring in the completion study. Table 5.4 lists the number of expected answers (out of 5) grouped by the type of HB words provided. Even though the number of expected answers is smaller, the results of this study are similar to the results reported in the sentence completion study. The number of expected answers increases together with the amount of biasing information provided. We obtained this outcome despite the fact that context words were presented in isolation rather than in a sentence.

The LME analysis reported in Table 5.5 treats the number of expected answers as the dependent variable, the number of HB words as the factor of the analysis that was contrast coded against the reference level *Zero*; `Participant` and `Item` were included as random intercepts and slopes. The *None* condition is significantly different from the *All* condition. It is also significantly different from the conditions that describe the effect of two HB words (*LocAct*, *ActObj*, *LocObj*). Among the conditions including one HB word (*Loc*, *Act*, *Obj*), only the *Obj* condition shows a significant difference from the reference level. As already discussed in the sentence completion study, an

Condition	Score
None	0.03 ± 0.02
Loc	0.28 ± 0.09
Act	0.33 ± 0.10
Obj	0.53 ± 0.11
LocAct	0.53 ± 0.14
ActObj	0.81 ± 0.14
LocObj	0.91 ± 0.14
All	1.05 ± 0.16

Table 5.4: Word completion study: Number of completions matching the target words (out of 5) with standard errors. The results are grouped by number of HB words and ordered by score.

Predictor	Coefficient
(Intercept)	0.03
Loc	0.25
Act	0.31
Obj	0.51^{**}
LocAct	0.51^{*}
ActObj	0.79^{**}
LocObj	0.88^{***}
All	1.03^{***}

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.5: Word completion study: LME Coefficients for the scores in Table 5.4.

HB object facilitates participants to identify the expected target word. The same effect emerges in both studies and indicates that the facilitation is mainly driven by the semantic similarity between the object and the target word. A post-hoc analysis shows that there is not significant difference between conditions describing the same amount of HB words ($p > .05$).

Figure 5.4 reports the number of expected answers averaged by the number of HB words. The pattern is similar to the one reported in the previous study (note the different scale in the two plots). After a context with only LB words (*Zero*) the averaged number of expected answers is 0.03 ($\pm .02$), it is 0.38 ($\pm .06$) in condition *One*, 0.75 ($\pm .08$) in condition *Two*, and 1.05 ($\pm .16$) in condition *Three*. Table 5.6 reports the LME analysis for these data. The number of expected answers was the dependent variable, the number of HB words was the factor of the analysis and was contrast coded against the reference level *Zero*; `Participant` and `Item` were included as random intercepts and slopes. Also in this study, the inclusion of at least one HB context word allows participants to perform significantly better than in the *Zero* condition. A Tukey

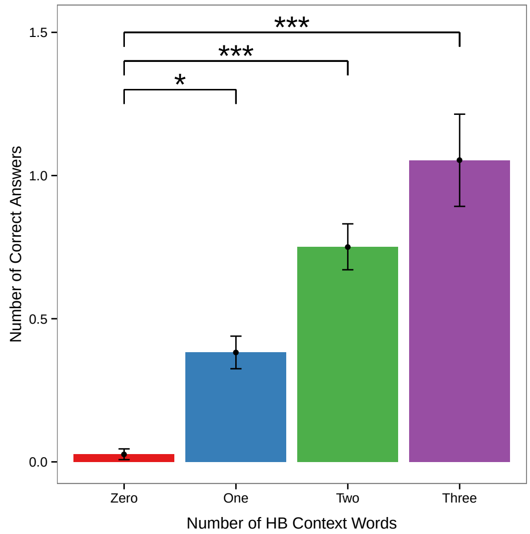


Figure 5.4: Word completion study: Number of expected answers averaged by the number of HB context words.

Predictor	Coefficient
(Intercept)	0.03
One	0.36^{*}
Two	0.72^{***}
Three	1.03^{***}

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.6: Word completion study: LME coefficients for the scores in Figure 5.4.

HSD test shows that the difference between conditions *Two* and *Three* does not reach significance (p=.082).

The aim of the word completion study was to exclude syntactic and word order effects from the results obtained in the sentence completion study. The results of the two studies (Figure 5.3 Figure 5.4) suggest a very similar trend: the inclusion of HB words in the context significantly facilitates the identification of the expected target word. Overall, the results reported so far show that the facilitation effect is mainly driven by semantic relations between context and target words. The presence of syntactic relations allows participants to produce a higher number of expected targets, but the differences between conditions remain constant. The comparison of the two plots in Figure 5.3 and Figure 5.4 shows that in the sentence completion study already one HB word is enough to produce a significant effect; while in the word completion study at least two HB words are required to produce the same effect. This is a first evidence in favour of the immediate activation hypothesis: a certain number of HB words are required to boost the pre-activation of the target word and make it more predictable. The different number of HB words required to produce an effect is related to the different complexity of the two experiments.

Procedure In the self-paced reading study, the 192 experimental sentences were distributed over eight lists of 24 items each according to a Latin square design. Twenty-seven fillers were added and the list randomised for each participant. Twenty yes/no questions about the sentence were also included.

Thirty-four native English speakers from the University of Edinburgh took part in the experiment after giving informed consent and were paid £5. Each subject saw one of the lists. We excluded one participant based on a low percentage of correct answers ($< 50\%$) and another participant with a RT averaging 2.5 standard deviations above the grand mean (as suggested in Hofmeister, 2011).

We used a moving-window self-paced reading procedure as discussed in Section 4.1. At the beginning of each trial, all the words in the sentence were masked with dashes and separated by spaces; participants had to press the space bar to uncover the next word and hide the previous one. To perform this experiment, we used the software package Linger[3] (version 2.94) on Apple computers.

5.2.2 Results

We analysed the RTs associated with the target word; they index the amount of effort required to process its meaning. Table 5.7 shows the mean RTs for the target word across the eight context conditions. The results indicate that RT decreases in proportion with the number of HB words in the context. Table 5.8 reports a LME model with log-transformed RT as the dependent variable. In the model, the contextual factors were contrast coded against the reference level `None`. `Participant` and `Item` were included as random intercepts, and as random slopes under all the contextual factors. The LME model shows that the conditions including two (*LocAct*, *LocObj*, *ActObj*) or three (*All*) HB words are statistically different from the *None* condition. Overall, after a HB context, participants spend less time reading the target word compared to a LB context. However, the difference in RT between *None* and the conditions with only one HB word (*Loc*, *Act*, *Obj*) does not reach significance. A post-hoc analysis shows no significant difference between the levels sharing the same amount of HB information ($p > .05$). In bold we highlight the conditions we include in the visual world experiment described in the next section.

In Figure 5.5 we report the RTs averaged by the number of HB context words. After three LB words it takes on average 390 ms to read the target word. The time

[3] http://tedlab.mit.edu/ dr/Linger

Condition	Reaction Times
None	390 ± 22
Act	369 ± 15
Loc	360 ± 18
Obj	349 ± 13
LocAct	350 ± 16
LocObj	350 ± 14
ActObj	329 ± 9
All	342 ± 11

Table 5.7: Reading time (in ms) with standard errors for the target word in the eight contextual conditions.

Predictor	Coefficient
(Intercept)	5.880^{***}
Act	-0.032
Loc	-0.068
Obj	-0.078
LocAct	-0.101^{*}
LocObj	-0.085^{*}
ActObj	-0.118^{**}
All	-0.094^{*}

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.8: LME coefficients for the RTs in Table 5.7.

decreases when the amount of HB information increases: 359 ms in condition *One*, 343 ms in condition *Two*, and 342 ms in condition *Three*. The LME analysis reported in Table 5.9 indicates that at least two HB words are required to significantly reduce the RTs compared to the reference level. A post-hoc analysis shows no significant differences between the other conditions ($p > .05$).

5.2.3 Discussion

In the self-paced reading experiment reported in this chapter, we looked at the relation between the number of HB context words and the RT on the target word. The hypotheses we started with predict two distinct outcomes: the incremental activation hypothesis predicts that the RT of the target word is reduced in proportion to the number of HB context words that are present. On the other hand, based on the immediate activation hypothesis, we expect that there is a threshold on the amount of contextual information that is required before an effect of context on RT occurs.

Descriptively, the RTs in Table 5.7 are compatible with incremental activation: each additional contextual word results in a further reduction in RT. However, the mixed model analyses in Table 5.9 shows significant differences between *Two* and *Zero* and

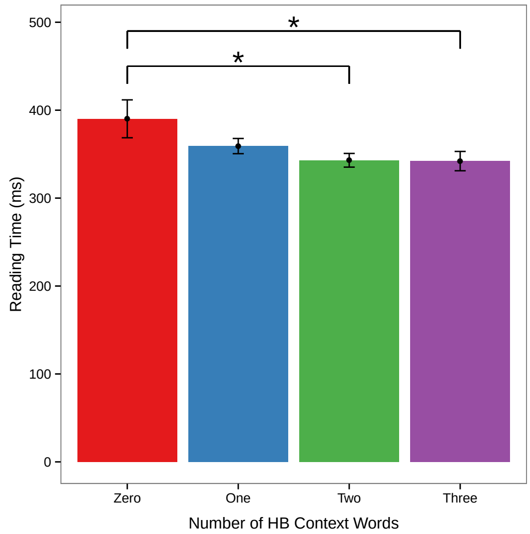


Figure 5.5: Plot of the reading times averaged by the number of HB context words.

Predictor	Coefficient
(Intercept)	5.888^{***}
One	-0.057
Two	-0.092^{*}
Three	-0.099^{*}

${}^{*}p < .05$, ${}^{**}p < .01$, ${}^{***}p < .001$

Table 5.9: LME coefficients for the RTs in Figure 5.5.

Three and *Zero*. The difference does not reach significance for condition *One*. This is a pattern that would be expected under the immediate activation hypothesis: the contextual threshold has been reached with two HB words (LocAct, LocObj, ActObj), and additional context words do not significantly pre-activate the target word any further.

These results allow us to identify a threshold in the amount of HB information that directly affects word processing; however they do not provide any description about the activation effect when each context word is encountered. We therefore performed a follow-up experiment in which we directly test our two hypotheses by measuring the amount of activation the target word receives during the processing of each context word.

5.3 Experiment 2: Contextual Incrementality over Time

The aim of this experiment is to test whether the activation of words happens gradually (more activation with every new context word, as predicted by the incremental activation hypothesis), or at once (the first context word triggers full activation, which then declines, as predicted by the immediate activation hypothesis). In a visual world study, we measure the amount of activation for the target word at each context word in terms of the proportion of looks received by the object corresponding to the target word.

5.3.1 Method

Materials The visual world paradigm requires both visual and linguistic stimuli. We used the same visual scenes as Huettig et al. (2006). For an overview of this study, see Section 3.3.

In our experiment, we used the same scenes of the original experiment: this allowed us to skip the norming process. In this way the association between the target word in the sentence and the corresponding four objects depicted on the screen was carefully considered. However, no norming study has been performed to control the relation between each context word and the images on the screen.

The sentence materials were the same as in Experiment 1, and the stimuli instantiated the same 24 target words. We chose four out of the eight original conditions in order to closely analyse the two hypotheses: *None*, *Loc*, *LocAct*, *All*. In this way, the amount of HB information provided is incremental. As an example, consider the visual stimulus in Figure 5.6, which corresponds to the sentences in (2),(3),(6), and in (9).

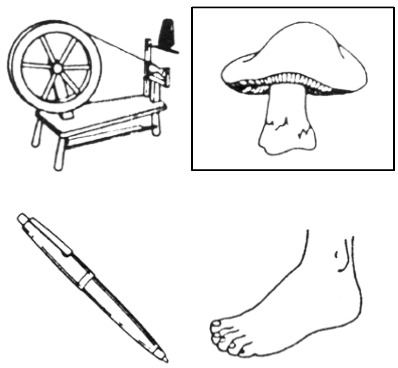


Figure 5.6: Example of the visual scene for the target word **mushroom** (the box is not shown to the participants).

Procedure The 96 sentences in the experiment were spoken by the speech synthesis system Festival (Clark, Richmond, & King, 2007) using an HMM voice (Roger), so as to reduce possible effects of prosody or speaker variation.

In order to counterbalance order or position effects, we rotated the four objects on the screen. The resulting 384 items were distributed over 32 lists of 24 items each according to a Latin square design. Twenty-five fillers were added and the list randomised for each participant. Nine yes/no questions about the sentence were also presented.

Thirty-four native English speakers from the University of Edinburgh took part in the experiment after giving informed consent and were paid £5. Each saw one of the lists.

Participants were seated in front of a 21” multi-scan monitor with a resolution of 1024 $\times$ 768 pixels and their eye movements were recorded using an EyeLink II head-mounted eye-tracker with a sampling rate of 500 Hz. Only the dominant eye was tracked. At the beginning of the experiment and after every ten trials, the eye-tracker was recalibrated using a nine-point randomised calibration. Before each trial, drift correction was performed. At the beginning of each trial the scene appeared on the screen, and the sentence began to play at the same time; the scene disappeared 1500 ms after the end of the sentence. The experiment lasted approximately 20 minutes.

Data Analysis The analysis is based on the proportion of fixations on the target object across experimental conditions. We excluded out-of-screen fixations and blinks from the analysis.

In order to analyse the effects exerted by a context word before and after its acoustic offset, we aligned the fixation probabilities at that point (0 ms). In order to exclude any overlap between two regions of analysis in the sentence we calculated the minimum amount of time between the onset and the offset of the context word (150 ms) and between the offset of the context word and the onset of the following one (400 ms for location, actor, and target word; 150 ms for object). The vertical line shows the offset of the context word, while the horizontal dotted line indicates the probability of randomly fixating on one of the four objects depicted on the screen (25% of total fixations).

For each context word we report an LME analysis of the results. As suggested by Barr (2008), the dependent variable of our models is the *empirical logit* of the fixation probability calculated for each bin. We used a bin size of 10 ms. To compare the effects produced by HB and LB contexts, we included three factors in contrast coding: each factor encodes the differences between the reference level `None` (coded as $-.5$) and one of the three other conditions (`Loc`, `LocAct`, `All`; coded as .5). The continuous factor `Time` shows variations over time. In order to identify the minimal model that best fits our data, we used the best-path forward model selection procedure (recommended by Barr et al. (2013) if the maximal model fails to converge). We report only the coefficients and the significance levels for the minimal model, i.e., we show only the main effects and the interactions included during the selection procedure. All models included `Participant` and `Item` as random intercepts, as well as random slopes for `Context` and `Time`.

5.3.2 Results

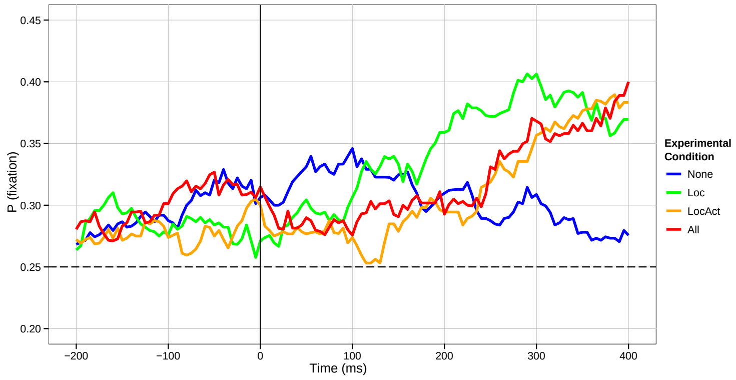
Predictor	Coefficient Location	Coefficient Actor	Coefficient Object	Coefficient Target Word
(Intercept)	-0.7548^{***}	-0.6764^{***}	-0.6215^{***}	-0.0138
Time	0.0005	0.0002^{***}	0.0003	0.0003^{***}
Loc	0.1019	-0.7303	0.1955	-0.1110
LocAct	-0.0388	0.4809	-0.0744	-0.2919
All	-	-0.1136	0.2204	-0.0849
Time:Loc	0.0008^{***}	-0.0011^{***}	-	0.0003^{**}
Time:LocAct	0.0003^{**}	-	-0.0024^{***}	-0.0007^{***}
Time:All	-	-	0.0011^{***}	-0.0004^{***}

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

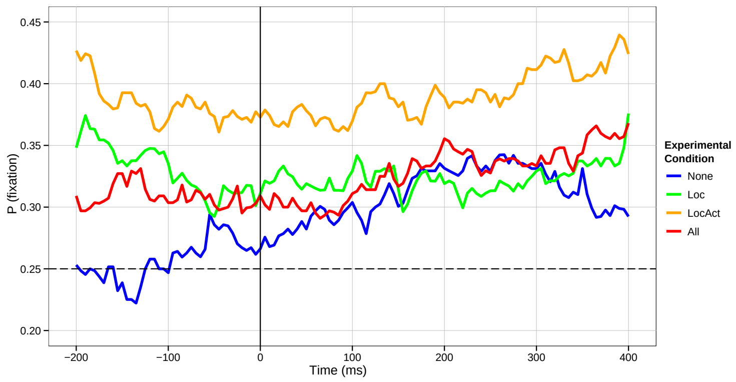
Table 5.10: LME coefficients for the data in Figures 5.7(a), 5.7(b), 5.7(c), 5.7(d). Empty cells indicate that the factor in question was not included during model selection.

Location Word The first context word we analyse is location. The plot in Figure 5.7(a) shows the probability of fixating the target object at this context word. Before its offset, it is already possible to identify a general effect produced by the presence of a location (both HB and LB), as the fixation probabilities are higher than random. However, specific effects appear only 100 ms after the offset of the context word. The *None* (low biasing) condition shows a decrease over time, while an increase in fixation probability is observed in the *Loc* and *LocAct* conditions (compared to *None*, the reference level), which corresponds to the significant interactions `Time:Loc` and `Time:LocAct` (see Table 5.10, column 1). A similar effect is visible for *All*, but fails to reach significance.

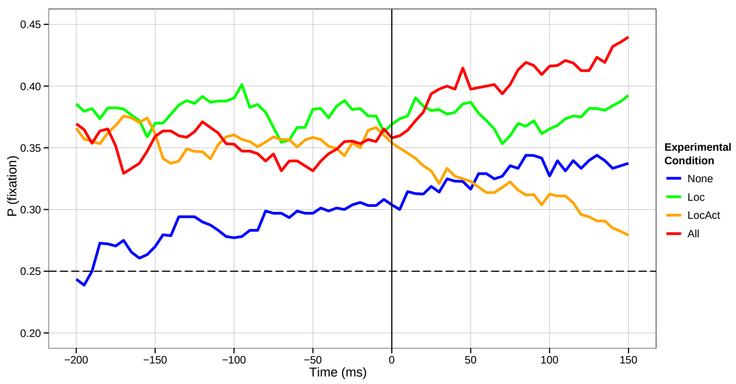
Actor Word The plot in Figure 5.7(b) shows the fixations at the word encoding an actor. In the *Loc* condition, participants tend to fixate less on target object (compared to *None*), an effect that is more evident before the offset of the word. This corresponds to a significant negative interaction `Time:Loc` in the LME (see Table 5.10, column 2).



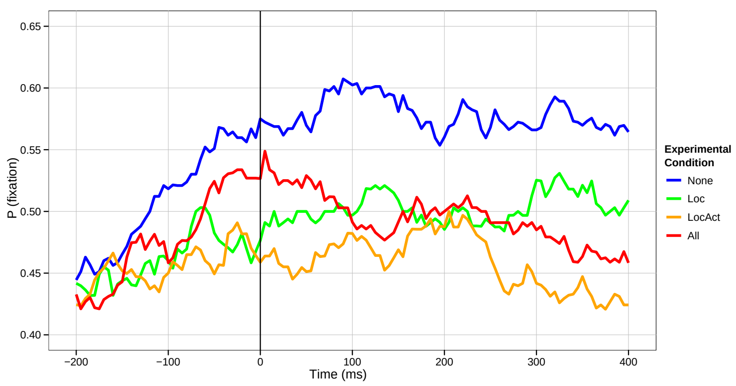
(a) *Location*: Target Fixation probability.



(b) *Actor*: Target Fixation probability.



(c) *Object*: Target Fixation probability.



(d) *Target Word*: Target Fixation probability.

Figure 5.7: Fixation probabilities aligned at the offset (0 ms) of the *context words*.

The plot also seems to indicate an overall higher level of fixations in the *LocAct* condition (compared to *All*, which is identical at this point in the sentence). However, this difference is not significant (no main effect or interaction involving *LocAct* in the LME).

Object Word The next context word analysed is the object of the sentence. Figure 5.7(c) shows that before the offset of the object, the HB conditions all show a higher overall fixation proportion compared to *None*. After the offset, the curves diverge: *All* shows a steeper increase than *None* (significant positive interaction `All:Time`, see Table 5.10, column 3). This is explained by the fact that *All* is the only condition with a HB object. The condition *LocAct*, shows a steep decrease, i.e., the significant negative interaction `LocAct:Time`, while *Loc* remains constant (no significant effects involving `Loc`).

Target Word Figure 5.7(d) shows the number of fixations at the point when participants hear the word associated with target object on the screen (note the different y-axis). At this point, global effects of different amounts of HB information across conditions should be visible. After the offset of the context word, there is an inverse relation between the amount of HB information and the slope of the curves in the *Loc*, *LocAct*, and *All* conditions. The more HB information is available, the sooner fixation proportions decrease. This is consistent with the pattern observed in the RTs of the target word in Experiment 1. A consistent pattern is also described in Experiment 5 (see Section 7.1).

On the other hand, fixation probability in the *None* condition increases, in particular after word offset, and remains high. The negative interactions `Time:LocAct` and `Time:All` (see Table 5.10, column 4) are consistent with this observation, indicating a significant decrease in fixations in *LocAct* and *All* compared to *None*. Furthermore, there is a significant positive interaction `Time:Loc`, indicating an increase in fixation probability in this condition compared to *None*.

5.3.3 Discussion

The aim of this experiment was to analyse the effect of incremental context information over time. The analysis of actor and object (see Figure 5.7(b) and Figure 5.7(c)) showed few interesting effects. The regions at which it was possible to identify a clear effect of contextual variability were location and target word. Location is the first context word participants are exposed to and it had a strong effect on driving their fixations towards the target word. This is in line with previous results of visual world studies on language comprehension (Altmann & Kamide, 1999), showing anticipatory eye movements towards a target as a result of predictive spoken language input. Less expected are the outcomes related to the target word area. We found that high-biasing contexts allow participants to identify and process the target object at an early stage: this effect is visible even at the first context word (location in our case). At the target word itself, we then fail to observe a sustained increase in fixations to the target object. The opposite pattern was observed in the low-biasing context: were we see no increase in fixations at the context words, but a sustained increase once the target word has been processed. In an HB context, the target word is contextually expected, and thus fixated less, while in the LB context, it is unexpected and thus fixated more.

One possible explanation for this pattern of results (i.e., a decrease in target fixations at the target word after a biasing context) is inhibition of return. This is a well-known effect in eye-movements, which manifests itself in a low probability of returning to a region once it has been fixated (Posner, Rafal, Choate, & Jonathan, 1985). Our failure to find an increase in fixations at the target region in the HB conditions could be due to inhibition of return, as the target had already been fixated at an earlier point in these conditions (i.e., during anticipatory processing while hearing biasing context words).

5.4 Experiment 3: Syntactic and Semantic Effects on Word Processing

In Experiment 1 and Experiment 2 we showed that not all context words affect word processing equally. Two HB words (location and actor) in the reading study and one HB word (location) in the eye-tracking study produce the major effect on the pre-activation of the meaning of the target word. The context words in these experiments were always in the same order: a location followed by an actor and by an object. For this reason we cannot clearly identify if the effect described is due to the order in the sentence (one HB word as first or second context word) or to the semantic type of the first context word (a location as first context word).

If the facilitation effect is due to the order in the sentence, the inclusion of an actor or an object as first context word should produce the same pattern described in the previous experiments. This outcome is supported by the results in the word completion study. This study showed that the order of the HB words in the context does not produce significant differences between conditions. Similarly, the post-hoc analysis in Experiment 1 indicates no differences between the three conditions with only one HB word (Loc, Act, Obj). However, if the facilitation effect is due to the semantic nature of the first context word, the inclusion of an actor or an object in position one should produce significant different results.

In order to test this hypothesis we performed a second self-paced reading experiment in which we manipulated the order of the context words in the sentence.

5.4.1 Method

Materials We used the same 24 target words and 72 contextual words (three for each target word) as in the previous experiment (cf. Section 5.3). Every context was the combination of three contextual words (a location, an actor, and an object). We swapped the order of the contextual words to have either a location, an actor, or an object as the first critical word. For example, the following sentences are those associated with the target word **mushroom** (in this example all the context words are HB words).

(10) *Location*: In the *forest* the *picker* was holding a *basket*, while the **mushrooms** were being picked carefully.

(11) *Actor*: The *picker* in the *forest* was holding a *basket*, while the **mushrooms** were being picked carefully.

(12) *Object*: A *basket* was held by the *picker* in the *forest*, while the **mushrooms** were being picked carefully.

As in the previous experiment, also the corresponding sentences with zero, one, or two high biasing words were included. In order to easily swap the different words in the context, we embedded the critical word in a subordinate phrase that follows the context sentence. The entire set of linguistic stimuli is reported in Appendix A.

Procedure The 288 experimental sentences were distributed over twelve lists of 24 items each according to a Latin square design. Twenty-seven fillers were added and the list randomised for each participant. Thirty multiple-choice questions after each sentence were also included.

Forty native English speakers from the University of Edinburgh took part in the experiment after giving informed consent and were paid £5. Each saw one of the lists. We excluded four participants based on a low percentage of correct answers ($< 85\%$).

5.4.2 Results

We analysed the RTs at the target word for each semantic category: a location, an actor, or an object as first context word. The analysis was performed using LME models with log-transformed reading time as the dependent variable. The `None` condition is the referent level for the analysis. `Participant` and `Item` are the intercept and random slopes of the model.

The first analysis we report is on the conditions in which a location is first context word. The context words follow the same order as in the previous self-paced reading experiment. Figure 5.8 shows the RTs averaged by condition. The results obtained do not replicate the outcome of the previous experiment. The LME reported in Table 5.11 shows no significant differences between conditions.

Figure 5.9 shows the RTs averaged by condition when an actor is the first context word. Similarly to the previous results, the LME analysis in Table 5.12 shows no significant differences between conditions.

Finally, Figure 5.10 reports the RTs averaged by condition when an object is the first context word. Again, no significant differences occur between conditions (see LME analysis in Table 5.13).

5.4.3 Discussion

The aim of this experiment was to test if the effect we showed in Experiment 1 and Experiment 2 was driven by the order or the semantic class of the three context words. Experiment 1 showed that at least two HB words are required to boost the activation of the target. Experiment 2 showed that the same effect occurs when the first context word (always a location) was a HB word. Taken together the results of these two experiment do not show if the effect is driven by the order or the semantic type of the context words. If the effect is driven simply by the order of the words (as we expect based on previous evidence) the same pattern should occur in all the results reported. However, if the semantic type is the reason for the outcome reported, we expect to see significant differences in the new results.

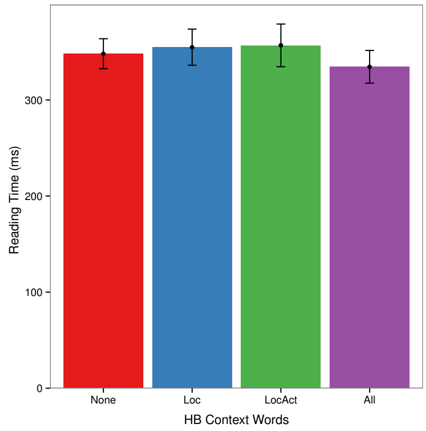


Figure 5.8: Location: Plot of the RTs averaged by type of HB context words.

Predictor	Coefficient
(Intercept)	12.7018***
Loc	−0.0005
LocAct	0.0023
All	−0.0382

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.11: Location: LME coefficients for the RTs in Table 5.8.

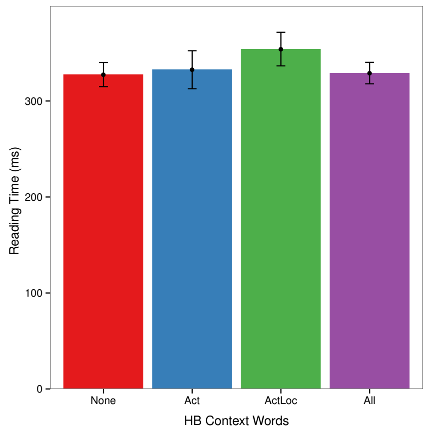


Figure 5.9: Actor: Plot of the RTs averaged by type of HB context words.

Predictor	Coefficient
(Intercept)	12.6523***
Act	−0.0009
ActLoc	0.0510
All	0.0116

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.12: Actor: LME coefficients for the RTs in Table 5.9.

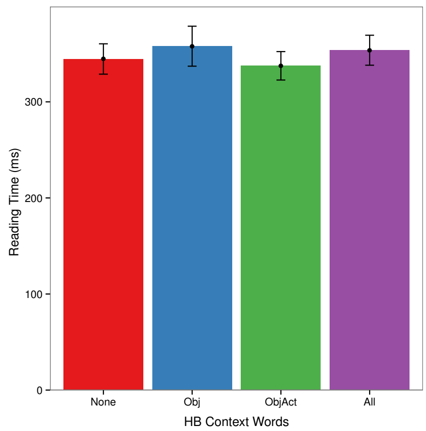


Figure 5.10: Object: Plot of the RTs averaged by type of HB context words.

Predictor	Coefficient
(Intercept)	12.6872***
Obj	0.0208
ObjAct	−0.2160
All	−0.0313

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.13: Object: LME coefficients for the RTs in Table 5.10.

The results reported show that there are no differences driven by the increasing amount of biasing words in the context. This null effect affects also the sentences that replicate the order of the context words used in the previous study (location, actor, object).

The main difference in the design of the two experiments stays in the syntactic structure of the linguistic stimuli. The stimuli used in this experiment include a subordinate clause that embeds the target word.

The use of the discourse connective “while” to introduce the subordinate clause affects the processing of the entire sentence. The use of this connective provides a very high number of plausible completions and this makes really hard for the participants to identify the expected target words. For this reason, it is clear that the stimuli that include a subordinate clause with so different possible completions are not suitable for our study.

In Chapter 8 we suggest several solutions to overcome this problem. In the next section we report 4 word association experiments. These studies allowed us to manipulate the order of the context words without been limited by syntactic constraints.

5.5 Experiment 4: Testing the Associations between Context and Target Words

In this section we report four association studies (4a-4d) that aim to shed light on the semantic relation between contextual words and the target word without syntactic constraints. The word completion study suggests that the order in which the three context words are provided does not affect the amount of expected target words that participants provide. In Experiment 3 (Section 5.4) we manipulated the order in which the three context words (a location, an actor, and an object) were appearing in the sentence. We were expecting to replicate the effect driven by increasing amounts of HB words described in Experiment 1 (Section 5.2). However, the study did not show any significant difference between experimental conditions. A possible reason for the null effect can be found in the syntactic structure of the new sentences used in the experiment.

To manipulate the order of the context words without any syntactic constraint, we performed a set of association studies in which we test the relation between context and target words in isolation. In Experiment 1 and Experiment 2 we analysed the effect of context on the target word. This effect is driven by at least two main factors:

the semantic relation of each context word with the target and the relation among the context words. The association studies make possible the analysis of these two factors independently. Moreover, the association scores collected in this experiment are used to evaluate the outcomes of a distributional semantic model in Chapter 6.

All the experiments were performed on Amazon Mechanical Turk. Subjects were required to rate how related two words are on a scale from one (not at all related) to five (very related). Subjects were all native speakers of English with an US account. They were payed $ 0.20 to produce 24 association scores and were allowed to complete only one hit from the same batch.

The LME analyses reported in this section have all the same structure. The association scores as the dependent variable, type or amount of HB words as the factors (the condition with the highest number of LB information is the reference level), `participant` and `item` as random slopes and intercepts. We also manipulated the order we presented the words (e.g. Context-Target, Target-Context): this was randomised and balanced over the entire experiment. We included this distinction as a factor in the model; however the analysis showed that the order of presentation has no effect either as main effect or in interaction with the different contextual conditions. For this reason, in the next sections we do not report the effect of order in the analyses.

Experiment 4a describes the associations between each contextual word and the target word; Experiment 4b analyses the associations between pairs of context words without including the target word; in Experiment 4c we consider the relation between triplets of context words without the target word while, in Experiment 4d, we include the target word.

5.5.1 Experiment 4a: Target-Context Associations

Aim The aim of this study is to investigate the semantic relation between each contextual word and the target word in isolation (e.g. forest-mushroom). Originally, this association study was performed as part of the pre-test analyses to assess the quality of the stimuli used in the reading and in the visual world study.

Method Sixty participants took part to the experiment. They were asked to rate the relation between each pair of context and target word. In total, 1,440 association scores were produced; each pair of context-target words was evaluated by ten subjects.

Bias	Type	Score
LB	location	2.74 ± 0.08
LB	actor	2.73 ± 0.07
LB	object	3.17 ± 0.09
LB		$\mathbf{2.87 \pm 0.04}$
HB	location	4.06 ± 0.07
HB	actor	4.10 ± 0.07
HB	object	4.32 ± 0.07
HB		$\mathbf{4.16 \pm 0.04}$

Table 5.14: Association scores with standard errors grouped by biasing effect and type of context word. The scores averaged by biasing effect are highlighted in bold.

Results Table 5.14 lists the association scores (out of 5) grouped by biasing effect (HB or LB) and type of context word (location, actor, or object). In bold we highlight the average scores for the HB and the LB words.

The HB words elicit higher association scores than the LB ones. A LME analysis shows that the difference between the two groups is statistically significant ($\beta_{HighBias} = 2.763$, $p < 0.001$). As already showed in the sentence and word completion studies, the object is more associated to the target word than the location and the actor but the difference between word types does not reach significance ($p > .05$).

Discussion The results show that HB words are more associated to the target than LB words. This study was a pre-test meant to evaluate the different semantic relation between HB and LB words and the target. The study confirms the intuition of the experimenters in differentiating LB and HB context words. It also shows that even with higher scores for the object, the semantic difference between word types (location, actor, and object) does not affect the associations between the context and the target word.

5.5.2 Experiment 4b: Context-Context Associations

Aim In this experiment we study the association between pairs of context words (e.g. forest-picker). The aim of the study is to understand how context words interact.

Method 144 participants took part to the experiment producing a total of 3,456 association scores. Each item was evaluated by twelve subjects. Participants were asked to judge the relation between each combination of LB and HB words for the following contextual pairs: loc − act, act − obj, obj − loc.

Results Table 5.15 reports the scores for the four possible combinations of HB and LB words. On average, two HB properties have the highest association score: 3.77 out of 5. On the other hand, two LB words obtain the lowest score: 2.81 out of 5. The associations between a high biasing and a low biasing word (*HB − LB* and *LB − HB*) are 2.93 and 2.96 out of 5. A LME analysis was performed (see Table 5.16). The only significant difference is between `LB−LB` and `HB−HB` words. We performed a post-hoc analysis to compare pairwise the different conditions. The `HB−HB` obtained significantly higher association scores than all the other conditions ($p < .001$).

Condition	Score
LB−LB	2.807 ± 0.044
HB−LB	2.927 ± 0.049
LB−HB	2.960 ± 0.049
HB−HB	3.773 ± 0.046

Table 5.15: Association scores with standard errors grouped by biasing effect (HB or LB biasing property) and ordered by score.

Predictor	Coefficient
(Intercept)	2.847^{***}
HB−LB	0.092
LB−HB	0.069
HB−HB	0.935^{***}

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.16: Coefficients for the LME for the association scores in Table 5.15.

Discussion The aim of this study was to analyse the semantic relation between pairs of context words. In this experiment we directly evaluated this relation without syntactic interference and the influence of the target word. The *HB−HB* condition obtained the highest scores, while the *LB−LB* condition the lowest ones. The mixed situations

(*HB−LB* and *LB−HB*) are negatively biased by the presence of the LB property in the pair. This result highlights an important element that has to be considered also for the interpretation of the results in the following association studies. The association of a HB word with a LB word is similar to the association of two LB words even though HB words have more specific meanings than LB words.

Overall, this study shows that the words that are highly related to the target word are also highly related among them. While, LB words (being more general words) appear also less strongly associated among them and also with the HB ones.

5.5.3 Experiment 4c: Multiple-Context Associations

Aim In Experiment 4b we analysed the relation between pairs of context words. In this study we test the association between three context words (e.g. forest-picker-basket).

Method 2,304 association scores were produced by 96 participants. Each item was scored by twelve subjects. The task required to evaluate the association between three context words. Eight combinations of LB and HB words were generated.

Results Table 5.17 reports the association scores grouped by the amount of HB information available and consequently ordered by score. The experimental conditions in bold are the same used in the visual world study (see Section 5.3). The scores increase together with the amount of high biasing information available except for condition *LB−LB−HB*. The LME analysis in Table 5.18 indicates that the association scores between two or three HB words are significantly higher than the reference level *LB-LB-LB*. The triplets with one HB object (*LB-LB-HB*) or one HB location (*HB-LB-LB*) do not differ significantly from the referent level while the condition with a HB actor (*LB-HB-LB*) shows a significant difference. A post-hoc analysis reveals that the difference between the conditions with the same amount of HB words is not significant ($p > .05$).

Figure 5.11 reports the results averaged by the number of HB words. The average association score for condition *Zero* (three LB words) is 2.83 (± 0.78), for condition *One* is 2.95 (± 0.05), for condition *Two* is 3.62 (± 0.5), and for condition *Three* is 4.32 (± 0.06). The LME analysis in Table 5.19 shows that at least two HB words are required to produce statistically higher scores than those in the reference level. As already discussed in the previous study, it is important to consider that the association

Condition	Score
LB−LB−LB	2.826 ± 0.078
LB−LB−HB	2.733 ± 0.077
HB−LB−LB	2.896 ± 0.082
LB−HB−LB	3.247 ± 0.084
HB−LB−HB	3.323 ± 0.081
LB−HB−HB	3.663 ± 0.077
HB−HB−LB	3.882 ± 0.076
HB−HB−HB	4.323 ± 0.060

Table 5.17: Association scores with standard errors between contextual triplets grouped by biasing effect (HB or LB biasing property) and ordered by score.

Predictor	Coefficient
(Intercept)	2.809^{***}
LB−LB−HB	−0.089
HB−LB−LB	0.065
LB−HB−LB	0.423^{*}
HB−LB−HB	0.523^{*}
LB−HB−HB	0.831^{***}
HB−HB−LB	1.080^{***}
HB−HB−HB	1.537^{***}

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.18: LME coefficients for the association scores in Table 5.17.

between one HB word and two LB words (as in condition *One*) is not different from condition *Zero*. A post-hoc test shows that the difference between all the conditions is highly significant ($p < .001$).

Discussion The aim of this study was to examine the semantic relation between three contextual words without the influence of the target word. The outcome is in line with the results obtained in the previous studies. Overall, an increasing amount of HB words produced a stronger association between words.

5.5.4 Experiment 4d: Target-Multiple-Context Associations

Aim Finally, we performed an association study where we included three contextual words and the target word (e.g. forest-picker-basket-mushroom).

Method 96 participants took part to the experiment and produced in total 2,304 judgements. Each item was scored by twelve subjects. The task required to evaluate

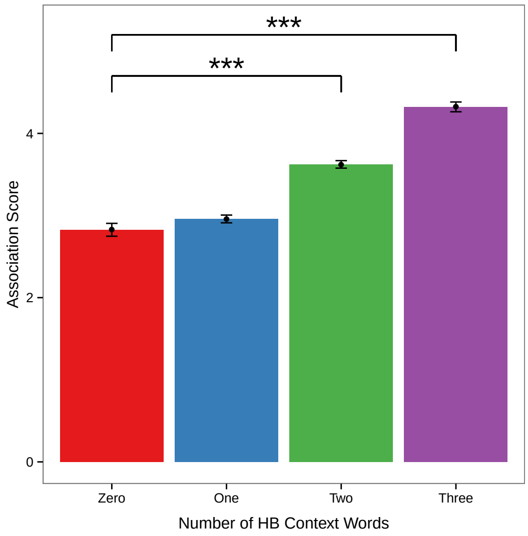


Figure 5.11: Association scores between triplets of words grouped by number of HB words.

Predictor	Coefficient
(Intercept)	2.788***
One	0.151
Two	0.834***
Three	1.660***

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.19: LME coefficients for the association scores in Figure 5.11.

the relation between four words (three contextual words and the target word). The design of the experiment was the same as in the previous section; the target word was always in the last position after the three contextual words. Participants were not aware of the difference between context and target words.

Results Table 5.20 reports the association scores ordered by condition and by score. The results follow the same trend as Experiment 4c with higher scores caused by the inclusion of the target word. The association scores increase along with the amount of HB information provided. The LME model in Table 5.21 shows that the differences between the reference level and all the other conditions are strongly significant. A post-hoc analysis shows no differences between triples with the same amount of HB information ($p > .05$).

Figure 5.12 reports the association scores averaged by the number of HB words. The inclusion of the target word in the study produces a significant increase of the scores in all the conditions. The difference between condition *Zero* and all the other conditions is highly significant. A post-hoc test confirms that also the difference between all the levels are highly significant ($p < .001$).

Discussion The increasing amount of HB information produces a general increase in the scores. The scores in this study are overall higher than those reported in Exper-

Condition	Score
LB−LB−LB−T	2.969 ± 0.077
LB−LB−HB−T	3.434 ± 0.078
HB−LB−LB−T	3.566 ± 0.078
LB−HB−LB−T	3.819 ± 0.072
HB−LB−HB−T	3.924 ± 0.067
LB−HB−HB−T	4.132 ± 0.058
HB−HB−LB−T	4.253 ± 0.059
HB−HB−HB−T	4.517 ± 0.048

Table 5.20: Association scores with standard errors between contextual triplets and the target (T) grouped by biasing effect (HB or LB property) and ordered by score.

Predictor	Coefficient
(Intercept)	2.939^{***}
LB−LB−HB−T	0.481^{***}
HB−LB−LB−T	0.613^{***}
LB−HB−LB−T	0.863^{***}
HB−LB−HB−T	0.996^{***}
LB−HB−HB−T	1.199^{***}
HB−HB−LB−T	1.320^{***}
HB−HB−HB−T	1.618^{***}

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.21: LME coefficients for the association scores in Table 5.20.

iment 4c, and the differences between the reference level and all the other conditions are highly significant (Table 5.22).

5.5.5 Association Studies: Discussion

In this section we reported four association studies in which we analysed the semantic relation among contextual words, and between context words and the target word. In these studies we were free to manipulate the order and the number of HB words without syntactic constraints. Experiment 4a was a pre-test performed to study the relation between HB and LB words and the target words. The results show that HB words are more related to the target than LB words. Experiment 4b analysed the relation between pairs of contextual words. In this way we excluded any semantic effect introduced by the target word. The results of this experiment show that words that are more related to the target word (HB words) are also more related to each other; the presence of a LB word is significantly reducing the final score. In Experiment 4c we analysed the associations between three contextual words: again the association scores increase

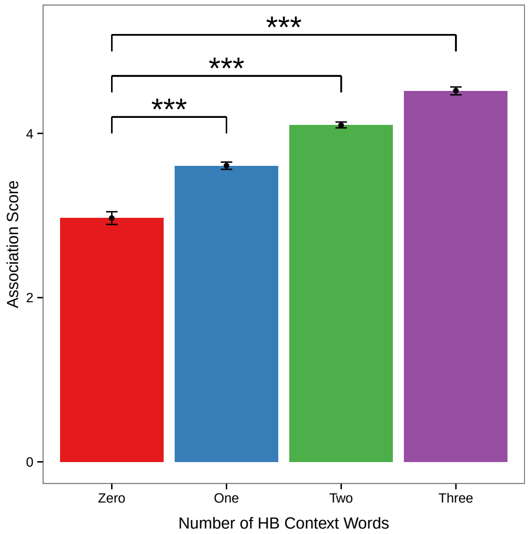


Figure 5.12: Association scores between three context words and the target averaged by the number of HB words.

Predictor	Coefficient
(Intercept)	2.913^{***}
One	0.678^{***}
Two	1.196^{***}
Three	1.636^{***}
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$	

Table 5.22: LME coefficients for the association scores in Figure 5.12.

along with the number of HB words. Finally, the inclusion of the target word in the study produces a general increase in the scores as shown in Experiment 4d.

The post-hoc analyses performed on the data in all the four experiments indicate that the different semantic classes of the context words (location, actor, object) do not significantly affect the resulting association scores.

Overall, the results of the experiments reported in this section support the outcomes of the previous studies. As the number of HB words increases, the semantic relations between these words become stronger (higher association scores). Moreover, they show that the order and the semantic class of the context words do not affect the final result. In Chapter 6 we use these association scores to test the capability of a distributional model in several classification and prediction tasks.

5.6 Conclusion

The aim of the studies reported in this chapter was to test how word processing is affected by increasing amounts of biasing context. In the introduction, we proposed two possible hypotheses on how context interacts with word processing: the *incremental activation hypothesis* and the *immediate activation hypothesis*. Both these hypotheses follow the main assumption of the predictive account (Section 2.2): context directly affects word processing and it makes it possible to pre-activate the meaning of the upcoming words in the sentence. The incremental activation hypothesis predicts that the degree of facilitation increases with the amount of context available. The *immediate activation hypothesis* predicts that once sufficient contextual support has accrued, no additional facilitation occurs.

In order to test these hypotheses and to clarify the relation between the amount of biasing information provided and the processing of the target word we performed six experiments: a sentence and a word completion study, two self-paced reading experiments, a visual world experiment and a series of association studies.

The main reason to perform the two completion studies (see Section (9)) was to test the overall quality of the materials used in the following experiments. In addition, they show the effect of context on word processing and, in particular, of the immediate activation hypothesis. Both the sentence completion study and the word completion study show that the inclusion of one HB word in the context is enough to significantly increase the number of expected target words produced by the participants. Moreover, they show that the inclusion of a HB word following another HB word does not produce any significant effect.

In the first self-paced reading experiment (Experiment 1, see Section 5.2) we investigated how an increasing amount of HB words affects the RT of the target word. The results reported show that two HB context words are required to boost the activation of the target word. When the activation takes place, no additional facilitation effect is produced on the RT of the target.

In the visual world paradigm experiment (Experiment 2, see Section 5.3) we directly controlled for the effect of the contextual information on the number of fixations towards the target object when the context was unfolding. This experiment shows an increased amount of looks to the target object during the processing of the first context word (location) and at the target word itself. The results indicate that a biasing context leads to an early recognition and processing of the target word. Moreover, the results point out an expectation effect: over time, the biasing contexts produced an expectation of the target word and this led to a reduced number of fixations to the target object when the corresponding target word occurred (potentially involving inhibition of return as the driving mechanism behind this decrease in fixation probability). The contribution of the study is to elucidate the time course of contextual integration. The results of Experiment 2 allow us to evaluate the two activation hypotheses. We found that when participants are exposed to a low-biasing context, we see an increase of fixations to the target word only when that word is mentioned. In the high-biasing conditions, on the other hand, this increase occurs already at the first context word, with no further increases at the second or third context word. Also when at the target word, only a small increase in fixations is observed. It seems that a single context word is sufficient to identify the target on the screen. Additional context exerts only a confirmatory effect.

The pattern of results in these two experiments is compatible with the immediate activation hypothesis: a certain amount of contextual information is sufficient to trigger word processing; additional contextual information does not produce an incremental increase in word activation. Comparing the outcomes of Experiment 1 and Experiment 2 we see a certain incongruity in the amount of HB information required to boost the activation of the target word. In the reading study at least 2 HB words are required in order to see a significant reduction in the RTs. While, in the visual world study the facilitation already takes place after the first HB word. A possible explanation for this difference in the results is that in the visual world paradigm experiment the visual information is always present and produces an higher facilitation effect on the activation of the target word. While during the reading task participants can only read one word at a time (without the option of reading again the already processed sentence): this makes the processing more difficult. In Chapter 7 we describe a blank screen paradigm experiment in which we reduce the facilitation effect produced by visual information during a language comprehension task.

In a second self-paced reading experiment (Experiment 3, see Section 5.4) we tested if the order in which the context words were presented or their semantic class were affecting the results reported in Experiment 1. Based on the results shown in the word completion study we expected to see that the effect was only driven by the order of the context words. If this expectation was valid, a HB location, actor or object as first HB word should produce the same facilitation effect on the processing of the target word. We designed the linguistic stimuli in order to have a location, an actor, or an object as the first context word. However, the introduction of a subordinate clause in the stimuli isolated the target word (in the subordinate clause) from the effect of context (in the main clause). In Chapter 8 we discuss possible solutions to overcome this problem.

In order to manipulate the order and the biasing effect of the context words without syntactic constraints, we performed four word association studies (Experiment 4, see Section 5.5). In these studies we asked participants to rate the relation between context words and between context and target words. Overall the results show that words that are highly related to the target word are also highly related to each other. Overall, the effect described is in line with the predictive account: it shows an increasing effect in the association scores based on the amount of HB words included.

In more theoretical terms, these results enhance our understanding of word representation. They indicate that the activation of a specific word takes place when the overlap between its internal structure and the information extracted from the context reach a certain critical level. The association studies have shown a strong relation between HB words. This result can be explained in terms of feature overlap theory: when context unfolds, it activates a certain amount of semantic properties. The upcoming words that match those features are pre-activated and processed more quickly when encountered. The high association between HB context words strongly restricts the set of possible upcoming words and makes the prediction task easier to perform. For this reason, when we increase the number of HB words in the context we do not see any different result because the pre-activation has been already taken place. As we discuss in Chapter 8, we expect to see a more incremental effect (as predicted by the incremental activation hypothesis) if the context words are highly related to the target but not highly related to each other. We provide more evidence about the feature overlap theory in the next Chapter.

Chapter 6

Modelling Contextual Effects on the Activation of Word Meaning

In the studies reported in the previous chapter we analysed the effect exerted by an increasing amount of biasing context on the activation and processing of word meaning. The results are consistent with the predictive account (see Section 2.2) with certain constraints. We showed that when the context is encountered, it affects the processing of upcoming words. However, a certain amount of contextual information is required to boost the activation process. When the expected word has been pre-activated, an increasing amount of context does not produce any additional significant effect. The experiments performed so far allowed us to study this process at the word level (e.g. how many words are required to facilitate the processing). As we discussed in Section 2.4, it is possible to describe the predictive account in terms of feature overlap. This theory suggests that when context unfolds, it activates a certain amount of semantic properties. The upcoming words that match those features are pre-activated and processed more quickly when encountered.

In this chapter we present five experiments in which we model several aspects of contextual incrementality using the bag-of-words distributional model developed by Mitchell (2011). As discussed in Section 3.2, bag-of-words distributional semantic models represent the meaning of a word as a multidimensional vector: each dimension of the vector corresponds to another word co-occurring with it in a corpus. Similarly to semantic properties, the vector dimensions describe specific aspects of a word that contribute to the meaning of that word. In this framework, two words are similar if they appear in similar contexts, and, consequently, if their vector dimensions overlap. Traditionally, the overlap between two words have been computed in terms of the geometrical distance between the word vectors. We will show that it is possible to describe the relation between low and high biasing words and the target word in terms of vector similarity scores.

First of all, we report the results of a classification task in which the model has to discriminate between low and high biasing context words. The classification is based on the similarity scores between context words and the target word. We show that the scores for the low biasing context words are significantly lower than those for the high biasing words. Similarly, we show that the model can predict the relation between context and target words as humans do in an association study (see Section 5.5.1). The similarity scores between context and target vectors positively correlate with the corresponding association scores between the same pairs of words.

In a second study we test the capability of the model in predicting the association scores between two context words (see Section 5.5.2), finding a significant positive correlation between human and model generated data.

In a third study we test the bag-of-words distributional semantic model in a further prediction task. Participants were successful in evaluating the associations between triplets of context words (see Section 5.5.3): the bag-of-words model produces similarity scores that positively correlate with these association scores.

The goal of the final study reported in this chapter is to show that distributional semantics provides a computational implementation of feature overlap theory. We show that it is possible to describe semantic features in terms of vector components. In order to do that, we model the accumulation of these features as the composition of the vectors of the context words. The similarity scores between the composed context vectors and the target words positively correlate with the reading times collected in Experiment 1 (see Section 5.2).

The chapter is structured as follows. Section 6.1 describes the distributional model we used to study the relation between context words and the target word in five different tasks. In Section 6.2 we describe the results of a classification task in which we test the model in discriminating low and high biasing words. In Section 6.3 we extend the analysis by using the model to predict human-association scores between contextual words and the target word. In Section 6.4 we report another study where we predict the association scores between pairs of context words. Section 6.5 describes a prediction task on association scores in which contextual words are represented as composed vectors. Finally, Section 6.6 reports a modelling study aimed to predict the reading times collected in Experiment 1.

Name	Description	Components
• condP	Conditional Probabilities	$p(c\|t) = \frac{freq_{ct}}{freq_t}$
• pmi	Point-wise Mutual Information	$log(\frac{freq_{ct} freq_{total}}{freq_c freq_t})$
• posPmi	Positive PMI	$max(0, log(\frac{freq_{ct} freq_{total}}{freq_c freq_t}))$
• ratiosOfP	Ratio of Probabilities	$\frac{p(c\|t)}{p(c)} = \frac{freq_{ct} freq_{total}}{freq_c freq_t}$

Table 6.1: List of association measures. $freq_{ct}$ is the frequency of the target t in the context c; $freq_t$ is the overall frequency of t; $freq_c$ is the overall frequency of c; and $freq_{total}$ is the total frequency of all the words (Source: Mitchell, 2011, p. 45).

6.1 The Model

For our study, we used a re-implementation (by Blacoe & Lapata, 2012) of the co-occurrence-based distributional model developed by Mitchell (2011). To weight the vector components, Mitchell used four different association measures: conditional probabilities (condP), point-wise mutual information (pmi), positive point-wise mutual information (posPmi), and ratio of probabilities (ratiosOfP). Table 6.1 shows how these measures are computed.

The original model was trained on the British National Corpus (BNC)[1], a corpus of around 100 million tokens. The linguistic information used to generate the vector representations was extracted from a contextual window of five words on both sides of the target word. A list of 770 stop-words was also provided.

As shown in Mitchell, the ratiosOfP model with vectors of 2000 components obtained the highest correlation scores both on a similarity task and on a synonymy task. Despite the small amount of parameter-settings and pre-processing required, this co-occurrence model reaches results similar to those from more complex models (as shown in Blacoe & Lapata, 2012).

Dealing with Infrequent Words Word frequency is a crucial issue in distributional semantics. Extremely frequent words cause an overestimation of the word similarities because they occur with almost every other word in the corpus. On the other hand,

[1]http://www.natcorp.ox.ac.uk

very low-frequent words do not appear enough in the corpus in order to be successfully modelled. This is one of the reasons why, nowadays, the majority of the studies uses a restricted amount of stimuli controlled in terms of their frequency (e.g. the WordSim353 test collection from Finkelstein et al., 2002).

Given the different nature of the study reported in this thesis, the linguistic stimuli were not extracted from one of those collections. As discussed in Section 5.2.1, some of our stimuli (among the high-biasing contextual words) are low frequency words with a very specific meaning (e.g. *hitches*); on the other hand, other stimuli (among the low-biasing words) are frequent, general words (e.g. *room*). Even though the extreme variability in the frequencies was not affecting human judgements (as shown in the norming studies on the stimuli for Experiment 1), it is a problem for the model. The original model trained on the BNC corpus did not find enough occurrences of some of the low frequency words in order to produce a satisfactory semantic representation of them. For this reason, we re-trained the model on a bigger corpus: the lemmatised and part-of-speech tagged version of ukWaK, an English corpus of two billion tokens extracted from the Web (Baroni, Bernardini, Ferraresi, & Zanchetta, 2009). The use of this corpus provided the full coverage of the experimental items.

We report the results obtained with each of the four association measures listed in Table 6.1 with eight different vector dimensions (from 1,000 up to 50,000 dimensions).

6.2 Study 1: Classifying High and Low Biasing Words

Aim The distributional semantic model (DSM) has to correctly discriminate between low biasing and high biasing contextual words. With this experiment we aim at answering a very general research question, related to the capability of DSMs to describe the biasing relation between context and target word in terms of vector similarity.

Task To construct the linguistic stimuli for Experiment 1 and Experiment 2 (see Section 5.2 and Section 5.3) we manually selected context words that are strongly (e.g. forest) or weakly (e.g. path) related to the target word (e.g. mushroom). A series of norming studies confirmed our intuition on the relation between each context word and the target. In this experiment, the model has to classify low and high biasing context words. If the semantic representation of the DSM successfully encodes this distinction, we expect the vector distance between a HB context word and its target word to be lower than the distance from a LB context word.

Method The model produced a vector representation for each of the 24 target words and the corresponding 144 context words used in Experiment 1. We computed the cosine distance between the vector of the target word and the vectors of the low and high biasing context words related to it.

Results Figure 6.1 reports the percentage of correct answers produced by the model: an item is considered correctly classified if the cosine distance between the target word and a low biasing word is higher than the distance between the same target word and the high biasing counterpart. Each line in the plot describes the performance of the model with a specific association measure (see Table 6.1) with varying vector dimensionality (from 1,000 up to 50,000 dimensions). The condP model (red line) overall obtains the worst results and it is not biased by the differences in vector dimensions. The pmi (green line) is the model that benefits more from the increased amount of vector components and it becomes stable at 30,000 dimensions. The ratiosOfP model (purple line) shows a negative effect produced by the increasing amount of dimensions. Finally, the posPmi (blue line) is the model that obtains the highest results in this classification task. The maximum score (90% of correct answers) occurs with vectors of 15,000 dimensions and remains stable when increasing the amount of dimensions.

We treat the cosine distances computed with the posPmi and with vector dimensionality set to 15,000 as the dependent variable of a LME analysis. To compare the difference between HB and LB context words we included the two factors in dummy coding: the reference level *High-Bias* (coded as 0) and the *Low-Bias* condition (coded as 1). The model included `Item` as random intercept and slope. The analysis shows that the cosine distance associated to low biasing words is significantly higher than the distance related to the high biasing ones ($\beta_{LowBias} = 0.123$, $p < 0.001$).

In Table 6.2 we report a qualitative analysis of the seven cases (10% of the total) when the cosine distances generated by the model are higher for the high biasing context than for the low biasing one. The column labelled "cosine difference" lists the numeric difference between the cosines of the two context words (high − low). For example, according to the model, *fridge* is .253 more similar to *room* than to *pub* in terms of their distributional relatedness. When looking at these results it is important to note that the cosine difference of only two properties is higher than .1 (*fridge* and *bus*). The smaller difference in the other pairs of words indicates that, in these cases, the model does not differentiate between low and high biasing context words. Only for *fridge* and *bus* there is an evident classification error. It is interesting that out of 7 wrong classifications 5 refer to a location. As discussed in McRae, Hare, Elman, and Todd (2005), locations are the less predictive contextual words; this is probably the reason why they are the words more difficult to be correctly classified.

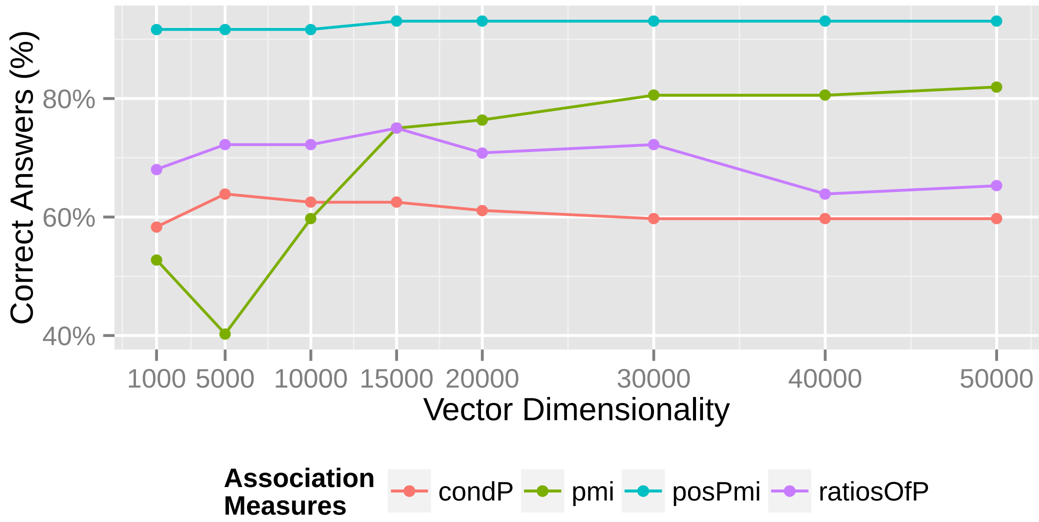


Figure 6.1: Percentage of correct answers when the model discriminates between low and high biasing contextual words with varying vector dimensionality. The distance between a low biasing context word and the target word has to be higher than the distance between a high biasing context word and the target word.

Discussion The aim of this experiment was to test the model in a classification task, based on the distinction between high biasing and low biasing words. Overall, this experiment shows that we can describe the relation between low/high biasing context words and the target word in terms of semantic similarity. High biasing words occur often together with the target word in similar situations described by similar linguistic contexts and the model is successful in capturing such similarities. The target word is more similar to high biasing context words than to low biasing context words because they share a higher amount of common dimensions. The posPmi model outperforms all the other models discriminating correctly the 90% of the high/low biasing pairs.

Target	High - Low Context	Context Type	Cosine Difference
fridge	pub - room	Location	0.253
bus	stop - road	Location	0.185
corn	straw - grass	Object	0.077
kettle	cafeteria - room	Location	0.051
mitten	scarf - jacket	Object	0.035
caterpillar	park - field	Location	0.022
coat	reception - dinner	Location	0.005

Table 6.2: Qualitative analysis of the classification errors generated by the model. The last column reports the differences between the cosines of the two context words (high-low).

6.3 Study 2: Predicting Target-Context Association Scores

Aim The aim of this study is to further investigate the relation between distributional similarity and the comparison between high and low frequency words. In particular, we test the capability of the DSM to predict association scores produced by experimental subjects.

Task In the association study described in Section 5.5.1, every pair of context - target word (e.g. forest-mushroom) was evaluated by 10 subjects on a scale from 1 (totally unrelated) to 5 (totally related). In total 1440 association scores were produced (24 target words, 3 context types, 2 biasing conditions). In this study, the model has to correctly predict these association scores.

Method We computed the cosine similarity for each of the context and target word pairs. In order to evaluate the performance of the model, we performed a correlation analysis between the human generated association scores and the cosine similarities computed by the model. We used the Spearman correlation test. Spearman *rho* is a non-parametric measure that does not assume a linear relation between association scores and model predictions; moreover, it is less prone to the presence of outliers in the data.

Results Figure 6.2 displays the correlation scores for the four models with varying vector dimensionality. Given the results of Study 1, there is a high similarity between the correlation scores reported here and the number of correct answers in Figure 6.1. Only the ratiosOfP model shows a different pattern obtaining the second best result in this task. Once again, the posPmi is the model with the highest coefficients for all the vector dimensions analysed.

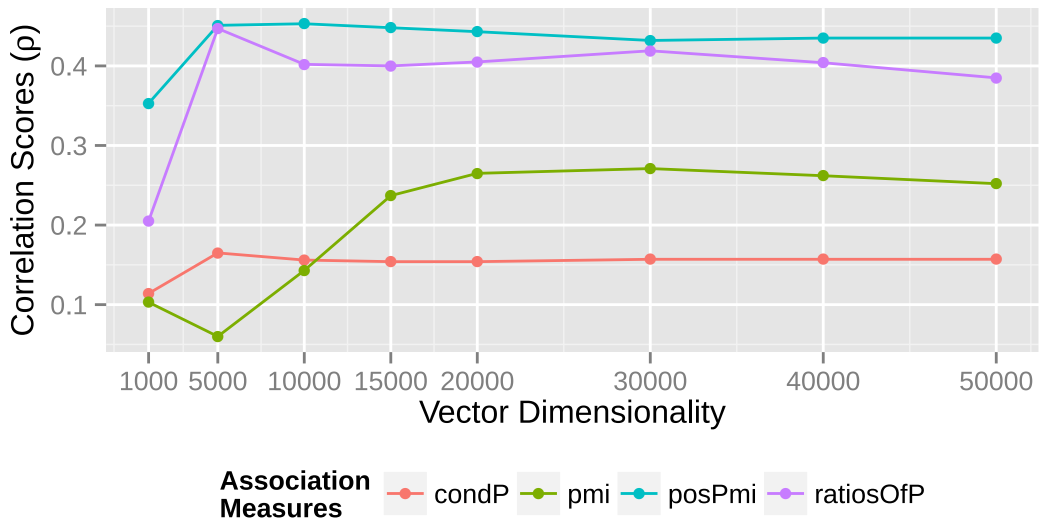


Figure 6.2: Spearman ρ coefficients for word similarities and human word-association scores with varying vector dimensionality. Each line shows the coefficients related to a different association measure. All the coefficients reported are statistically significant ($p < .001$).

We performed a LME analysis treating the association scores as the dependent variable, the cosine similarities (dim=10,000) as the continuous factor, and `Subject` and `Item` as random intercepts and slopes. The model indicates that the positive relation between word similarity and association scores is statistically significant ($\beta_{Cosine} = 5.020$, $p < 0.001$).

Discussion In this study we tested the DSM on the prediction of human-association scores between each contextual word and the target word. The experimental items involved in this modelling experiment are the same of Study 1. The perspective, however, is different because the type of evaluation performed in the association study in

Section 5.5.1 and modelled here was more fine grained: participants were not only categorising context words based on the high/low biasing distinction, but they were scoring the strength of this relation. Overall the pattern of the correlation scores is similar to the pattern of the percentage of correct answers reported in the previous study. Let us now briefly discuss the best parameter values in this task. The posPmi is again the model with the highest correlation scores for all the vector dimensions and the scores remain stable when increasing the vector components. The posPmi model generates word similarities that are comparable to human association scores (high association scores correspond to high cosine similarity). Surprisingly, both in Study 1 and 2, the posPmi outperformed the ratiosOfP model that, on average, obtained higher results in a previous synonyms identification task and in a similarity judgement task (Mitchell, 2011). A possible explanation for those results can be found in Mitchell (2011, p. 44): the substitution of negative values with zeros in the posPmi model (cf. Table 6.1), makes this association measure stronger in dealing with sparsity and low frequency words. A difference to the results achieved in Study 1 is the fact that the ratiosOfP model is performing definitely better in this task than in the previous one. This outcome suggests that this model, although producing an overall higher number of errors in the classification task, can precisely predict human-association scores.

The results achieved in Study 1 and Study 2 allowed us to identify a general trend for posPmi to be the best performing association measure. For this reason, in the follow up studies described in the next sections we only report the results for the posPmi model.

6.4 Study 3: Predicting Context-Context Association Scores

Aim The model is evaluated in the task of predicting human generated association scores between two contextual words (e.g. forest-picker). From the corresponding association study (see Section 5.5.2) it emerged that context words that are highly related to the target are also highly related to each other. We therefor expect the semantic similarity between two high biasing context words to be higher than the similarity between two low biasing context words.

Task In the association study described in Section 5.5.2, participants had to evaluate on a scale from 1 (not related) to 5 (completely related) the association between two contextual words (both high and low biasing words). A total of 3,456 association scores were produced: each pair of properties was evaluated by 24 participants. The model has to correctly predict these associations.

Method We computed the cosine similarity between pairs of context words. Based on the results of the previous two studies, we report only the performance of the system with the posPmi as the association measure.

Results Figure 6.3 reports the correlation scores with varying vector dimensionality between the human-association scores and the predicted similarities. The highest correlation score is obtained at 10,000 dimensions; after that there is a slight decrease in the scores. A LME analysis was performed. The association scores at 10,000 dimensions are the dependent variable, the cosine similarities the continuous factor, `subject` and `item` the random slopes and intercepts. The model shows a significant positive relation between word similarity and association scores ($\beta_{Cosine} = 6.431$, $p < 0.001$).

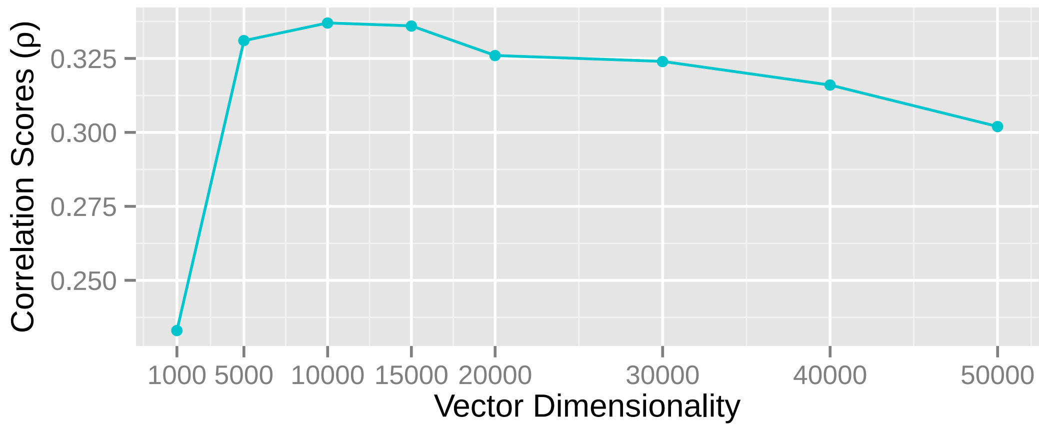


Figure 6.3: Spearman ρ coefficients for word similarities and human word-association scores, with varying vector dimensionality. The line shows the coefficients related to the posPmi association measure. All the coefficients reported are statistically significant ($p < .001$).

Discussion The results of this task point out that the model can successfully predict human association scores between context words. The analysis shows a general positive correlation between the two measures, with the highest score at dimensionality 10,000. The significant correlation between these two measures shows that the model can capture the relation between context words replicating the results of the association study. Words that are highly related to the target word are also strongly associated to each other because they occur in similar contexts.

6.5 Study 4: Predicting Target-Multiple Context Association Scores

Aim In this study we test the model in the task of correctly predicting the association between triplets of context words (e.g. forest-picker-basket) and the target word (e.g. mushroom). We represent the overall semantic meaning of contextual words by using vector combination.

Task In Section 5.5.3 we described a word association study in which participants were asked to judge the association between triplets of contextual words on a scale from 1 to 5. In Section 5.5.4 we performed the same experiment including the target word in the comparison. In each association study, a total of 2,304 human generated association scores were produced (each item was evaluated by 16 participants). We are now testing the model on these two prediction tasks.

Method In order to compute the cosine similarity between a target word and the contextual words, we produced a single contextual vector by combining the vectors of the single context words. Following Mitchell (2011), we test two possible composition methods: addition and point-wise multiplication.

Considering three context words w_1, w_2, and w_3, we computed the composed vector **context** by addition:

$\mathbf{context} = \mathbf{w\_1} + \mathbf{w\_2} + \mathbf{w\_3}$

or by point-wise multiplication:

$\mathbf{context} = \mathbf{w\_1} \odot \mathbf{w\_2} \odot \mathbf{w\_3}$

We computed the cosine similarity between the resulting **context** vector and the vector representing the target word. In both the experiments we used the same output of the model. The difference was in the scores produced by humans. In one study, participants were asked to evaluate the similarity among three contextual words alone (forest-picker-basket), in the other study, also the target word was presented (forest-picker-basket-mushroom).

Results Figure 6.4 and Figure 6.5 report the correlation coefficients between cosine similarities and human association scores with varying vector dimensionality. Each line describes one compositional method: addition (red line) or multiplication (blue line).

Figure 6.4 shows that the use of point-wise multiplication yields the best results in the prediction of human-association scores. After a peak at dimension 10,000, performance stabilises at dimension 30,000. According to a LME analysis with association scores as the dependent variable, cosine similarity as the continuous factor and `subject` and `item` as random intercepts and slopes, there is a strongly significant positive relation between association scores and cosine similarity at dimension 30,000 ($\beta_{Cosine} = 9.894$, $p < 0.001$).

Figure 6.5 shows a less defined pattern. Even though the multiplication model (blue line) obtained on average the highest results again, the difference between the two curves is not really considerable and there is a negative peak when the vector dimension is 20,000. Moreover, those scores were overall lower than those in the previous analysis. The LME analysis at dimension 40,000 shows a strong positive relation between associations and similarities ($\beta_{Cosine} = 8.449$, $p < 0.001$).

Discussion In this task, the model had to predict the human generated association scores between triplets of contextual words (as reported in Section 5.5.3). Similarly it had to predict the association scores between quadruplets composed by the three contextual words and the target word (as reported in Section 5.5.4). We combined contextual words in two ways: addition and point-wise multiplication of the vectors of the context words. As shown already in Mitchell and Lapata (2008), vector multiplication produces better results in these tasks. The dimensions that are not shared between all the vectors involved in the combination are set to zero. In this way, only the shared information stays active and contributes to the overall representation of the combined meaning. Overall, the results of this experiment show that point-wise vector multiplication is a good way of combining the meanings of context words. The similarity between the combined context vector and the target word vector successfully describes the relation between context words and the target word as shown by the association studies.

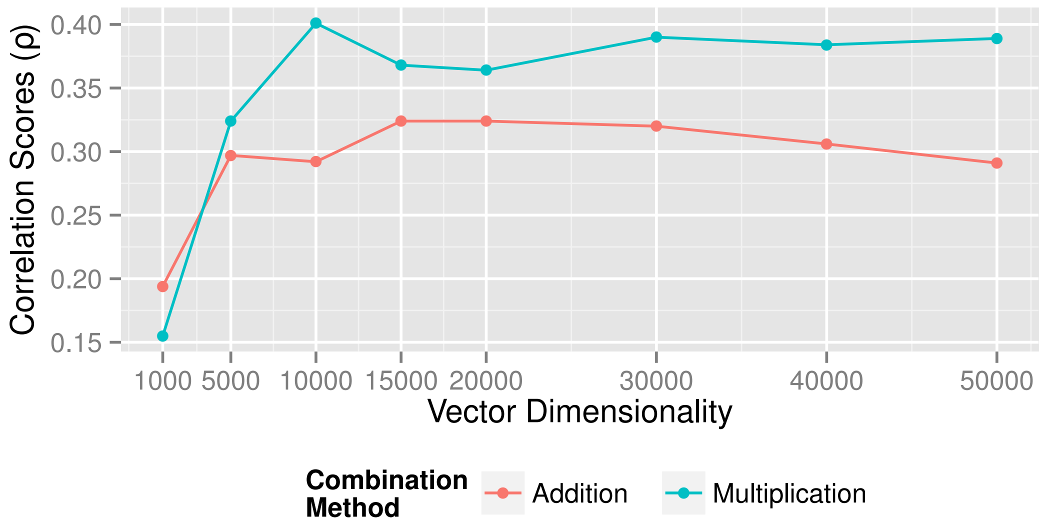


Figure 6.4: Spearman ρ coefficients between cosine similarities and human-association scores (only contextual words), with varying vector dimensionality. Each line describes a different composition method (addition and point-wise multiplication of the contextual vectors). All the coefficients reported are statistically significant ($p < .001$).

6.6 Study 5: Predicting Reading Times

Aim In the previous tasks we showed that the bag-of-words distributional model can significantly predict human generated association scores. In this study we test the model on the prediction of the reading times collected in Experiment 1. In Section 5.2.2 we showed that the amount of time required to read and process the target word decreases (even though not linearly) when increasing the amount of biasing words in the context. Similarly, the distance between the context vector (composed by the vectors of the three context words) and the vector of the target word should decrease when combining the vectors of high biasing words. Vice-versa, when combining the vectors of low biasing words, the cosine distance should increase.

Task In the self-paced reading experiment reported in Section 5.2.1, we analysed the incremental effect of context on word meaning. We averaged the RTs based on the number of HB words available (from zero up to three) as follows:

- **Zero**: no high biasing words;
- **One**: only one high biasing word;

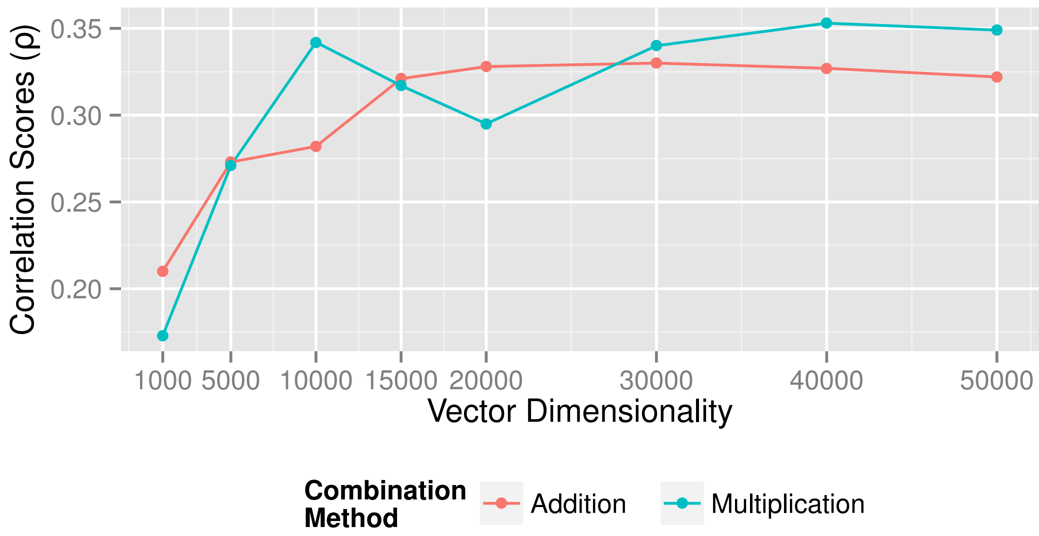


Figure 6.5: Spearman ρ coefficients between cosine similarities and human-association scores (including both target and contextual words), with varying vector dimensionality. Each line describes a different composition method (addition and point-wise multiplication of the contextual vectors). All the coefficients reported are statistically significant ($p < .001$).

- **Two**: two high biasing words;
- **Three**: three high biasing words.

In this study we compute the distance between the context vectors and the target vectors and we average the resulting cosines in the same way as we averaged the RTs.

Method As in Study 4, we combined the context vectors using point-wise multiplication. We computed the cosine distance between the resulting context vector and the vector of the target word.

Results Table 6.3 reports the average cosine distance per condition. It shows that the increasing amount of biasing context produces a reduction in the distance between the context vector and the target word vector. Table 6.4 reports the LME coefficients for these data. The model has the cosine distance as the dependent variable, the contextual condition as main factor (dummy coded with *Zero* condition as reference level) and `Item` as random slope and intercept. The model shows a significant difference between

Condition	Cosine Distance
Zero	0.991±0.005
One	0.975±0.008
Two	0.907±0.015
Three	0.835±0.016

Table 6.3: Average cosine distance with standard errors in the 4 conditions.

Predictor	Coefficient
(Intercept)	0.991^{***}
One	-0.015
Two	-0.084^{***}
Three	-0.156^{***}

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 6.4: LME coefficients for data in Table 6.3.

condition *Zero* and *Two* and condition *Zero* and *Three*. The coefficient increases in relation to the amount of biasing information provided.

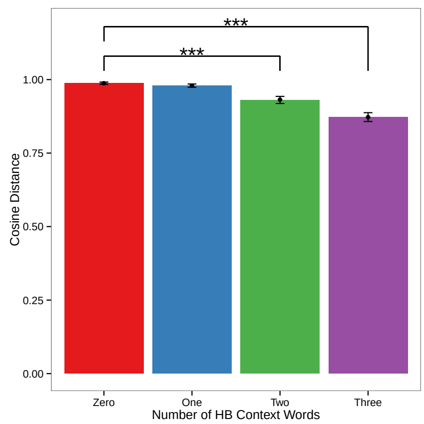


Figure 6.6: Plot of the cosine distance averaged by the number of HB words (cf. Table 6.3).

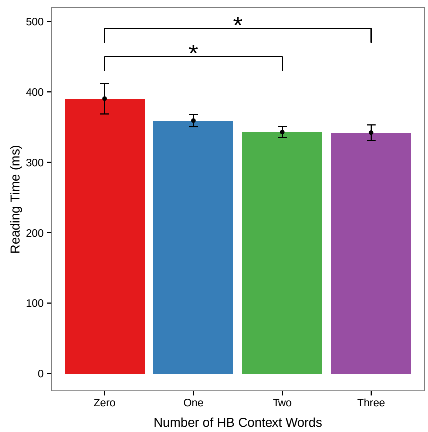


Figure 6.7: Plot of the reading times of Experiment 1 averaged by the number of HB context words (see Section 5.2.2).

Discussion Figure 6.6 and Figure 6.7 allow a graphical comparison of the trends in the cosine similarity study and in the reading study. The modelling study shows higher differences between conditions than those in the reading times produced by humans. Similar results have been described in the priming literature where LSA was predicting a stronger effect than the one observed with humans (Hare, Jordan, Thomson, Kelly, & McRae, 2009; Jones, Kintsch, & Mewhort, 2006). Overall, it is possible to identify a common trend: the need for a certain amount (two or more HB words) of contextual information in order to boost the activation process.

6.7 Conclusion

The aim of the experiments described in this chapter was to test a distributional semantic model in a classification study and in four prediction studies. The model successfully captured semantic relations similar to those emerged in the behavioural studies reported in Chapter 5. This outcome supports the cognitive validity of the model in describing the relations between words in terms of semantic similarity.

In Study 1 we evaluated the relation between the target word and low and high biasing context words. Overall, the model successfully discriminates between low and high biasing context words producing higher semantic similarity scores for the latter. As shown in Study 2, these scores correlate significantly with human generated association scores. Study 3 and Study 4 demonstrated that it is possible to model the association between contextual words and the target word combining the context word vectors. The combination of context words by point-wise vector multiplication generated the highest correlation scores. Finally, Study 5 provided support for feature overlap theory by showing that contextual facilitation increases with the number of highly biasing context words. We demonstrated that the accumulation of semantic features can be modelled as the composition of the distributional vectors of the context words. Distributional semantics therefore provides a computational implementation of feature overlap theory, with semantic features represented as vectors components (i.e., word co-occurrences).

Overall, we showed that the distributional semantic model can successfully capture various aspects of contextual effects on word processing. Based on the positive significant correlations between corpus based modelling and human judgements, we can conclude that the facilitatory effect produced by the context on the activation and processing of word meaning is driven by semantic relations between context and target words.

In this chapter we tested the validity of bag-of words DSMs as a representation of the relation between context and target words. In the next chapter we use Strudel, a different distributional semantic model, to directly produce the contextual words we embed in a new set of linguistic stimuli. As discussed in Section 3.2.2, Strudel generates vector representations that have feature-like properties as dimensions (Baroni et al., 2010). For this reason, the context words extracted by the model can be directly used as properties related to the target word. The effect of these properties on the activation of word meaning will be test in a visual world experiment.

Chapter 7

Contextual Effects on Semantically Similar Words

In this chapter we study the effect of context on pairs of semantically similar words and we analyse how semantic similarity relations between context and target words facilitate the pre-activation of the meaning of the target word.

Federmeier and Kutas (1999) demonstrated that when the expectations of an upcoming word are not satisfied, the semantic similarity between the expected and the provided word systematically reduces the N400 effect (see Section 2.4). Similarly, Huettig et al. (2006) analysed semantic similarity in a visual world paradigm (see Section 3.3). The authors showed that a word can drive eye-movements towards a semantically related object depicted on the screen. Both these experiments have described a strong effect of semantic similarity when accessing word meaning.

The novelty of our study consists in the manipulation of the amount of similarity between context and target words. In order to control for this relation we used a distributional semantic model to identify the best words to include in the context.

As discussed in Section 3.2.2, the main limitation in the use of traditional distributional semantic models is that their dimensions are not directly interpretable as semantic properties describing the internal structure of word meaning (Murphy, 2002). We reported several models that address this problem. In this thesis we decided to use Strudel (STRUctured Dimension Extraction and Labeling; Baroni et al., 2010). Strudel represents word meaning in terms of weighted interpretable typed properties and it extracts properties also of words that have not been already classified in an existing norming collection. The advantage of using these properties in our study consists in the stronger semantic relation that occurs between the properties and the word they describe. As the context unfolds, an increasing amount of these properties will overlap with the target word properties and facilitating its pre-activation.

The experiments also test the cognitive plausibility of distributional semantic models in generating properties that can be used as experimental stimuli.

Based on the outcomes of the studies reported in the previous chapters we expect to see a facilitation effect driven by semantic similar words on the processing of the target word. The higher the overlap between the semantic properties of the two words, the stronger the effect generated.

To test our expectations we perform a visual world paradigm experiment (Experiment 5 in section 7.1). A blank screen paradigm experiment (Experiment 6; section 7.2) produces results partially compatible with those in Experiment 5. The design of this experiment allows us to reduce the amount of visual information available during the task.

Experiment 5 appears in Frassinelli and Keller (2012).

7.1 Experiment 5: Testing Contextual Effects on Semantically Similar Words

As reported in Section 3.3, Huettig et al. (2006) used a visual world paradigm experiment to test the cognitive plausibility of “models of high-dimensional semantic space”. They used a list of 26 target/competitor pairs of semantically related but not strongly associated words. In every pair, one of the words corresponded to a target object depicted in a visual scene (the target word); the other one (the competitor word) was semantically related to the depicted object. For every pair of words, a spoken sentence was recorded that contained either the target or the competitor. Huettig et al. focused on the effect of hearing the target vs. the competitor as critical word. For this reason, the context sentences they used were neutral, providing background information that did not bias the participants towards either the target or the competitor. One of their contexts is given in (1) as an example.

(1) At first, the man laughed loudly, but then he saw the elephant (target)/alligator (competitor) and understood that it was dangerous.

The crucial manipulation in our experiment, however, concerns the context sentence. We run Huettig et al.’s neutral context as a baseline condition, but we add two context conditions: a context containing properties associated with the target, and a context containing properties associated with the competitor. These context sentences were constructed using three properties produced by the distributional model Strudel (Baroni et al., 2010) (see section 3.2.2). The model was trained on the lemmatized and part-of-speech tagged version of ukWaK (Baroni et al., 2009).

7.1.1 Method

This experiment aims to establish the effect that context has on the processing of semantically related words and to test the cognitive plausibility of using a distributional model to construct this context.

Huettig et al. used a neutral context and found that participants are more likely to fixate a target object when they hear its name, but they also show an increased fixation probability for the name of a semantically associated object. We expect this effect to be modulated by context. More specifically, when the properties associated with the target are processed, they should build up an expectation for the target, and as a consequence, there should be more fixations on the target object when the target word is spoken, compared to the neutral context condition. When the same properties associated to the target occur before a competitor word which is distinct from the target, but semantically related (as in Huettig et al.'s design), this effect should be attenuated but still present.

The outcomes of this experiment are relevant to confirm the key evidence provided in the previous studies: contextual information has an effect on word processing that is directly connected to the semantic relation that exits between a specific context and a word.

7.1.1.1 Materials

As in Experiment 2 (see Section 5.3), we used visual scenes that consisted of four black and white line drawings extracted from the Snodgrass and Vanderwart (1980) collection (already normed by Huettig et al., 2006): one target object and three distractors randomly arranged in four quadrants. The neutral context sentences were the same linguistic materials as Huettig. We added to this two context conditions: one for the target word, and one for the competitor. For each of the 52 words (26 competitor/target pairs) in the Huettig et al. materials, we extracted from the output of Strudel the first 20 semantic properties (nouns, verbs, and adjectives) ordered according to

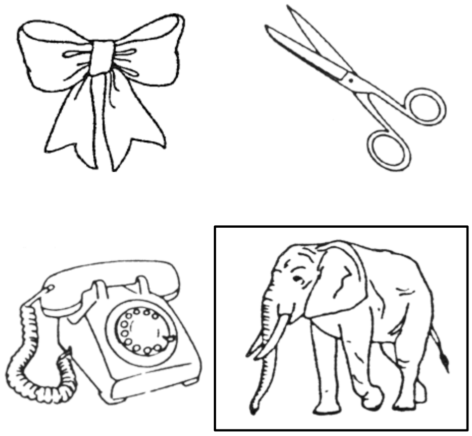


Figure 7.1: Example scene for the pair elephant (target)/alligator (competitor). The box highlights the target object (not shown to participants).

their log-likelihood ratio. We constructed a context sentence for each word using three of these properties (excluding those associated with words that are part of the same target and competitor pair). In this way we could manipulate the amount of semantic similarity between the entire context (composed by the properties generated for the target/competitor word or by neutral information) and the target/competitor word.

The context sentences have a standard pattern: a temporal subordinate clause introducing the situation followed by the main clause. The target word is embedded at the end of the main clause and followed by an adverb (which serves as a spill-over region for the analysis). Given the restricted pool of possible model-generated properties, the structure of the linguistic stimuli used here are less structured than the ones used in the previous experiments. As an example, Figure 7.1 depicts the scene associated with the pair elephant (target)/alligator (competitor). The six sentences associated with this scene are:

(2) `Neutral Context - Target Word`: At first, the man laughed loudly, but then he saw the **elephant** and understood that it was dangerous.

(3) `Neutral Context - Competitor Word`: At first, the man laughed loudly, but then he saw the **alligator** and understood that it was dangerous.

(4) `Target Context - Target Word`: While the man was crossing the *jungle*, he saw a *poacher capturing* an **elephant** ferociously.

(5) `Target Context - Competitor Word`: While the man was crossing the *jungle*, he saw a *poacher capturing* an **alligator** ferociously.

(6) `Competitor Context - Target Word`: While the man was crossing the *swamp*, he saw a hippo *attacking* a *gigantic* **elephant** ferociously.

(7) `Competitor Context - Competitor Word`: While the man was crossing the *swamp*, he saw a hippo *attacking* a *gigantic* **alligator** ferociously.

The critical word is given in **bold**; the properties are in *italics*.

Norming Studies The quality of the materials was evaluated in two norming studies performed using Amazon Mechanical Turk.

Sentence Plausibility Study In a sentence plausibility judgement task, 32 native English speakers rated the sentences on a scale from 1 (completely implausible) to 7 (completely plausible). The mean rating for the word in the sentence with the corresponding properties (e.g. sentences (4) and (7)) was 5.97 ($SE = 0.07$) and in the opposite sentence, it was 4.14 ($SD = 0.11$); the opposite sentences were created by swapping the critical words across conditions (target for competitor and vice-versa; e.g. sentences (5) and (6)). The stimuli with an average score below 3.5 have been discarded and substituted by better candidates. A LME analysis with `Item` and `Participant` as random slopes and intercepts showed a strong difference between the two conditions ($\beta_{Target} = 1.83, p < .001$).

Sentence Completion Study In a sentence completion task, we removed the critical words from the sentences and asked 21 participants to complete each of the 52 sentences (two groups of 36 sentences) by typing the most plausible noun. After a process of synonym reduction, we counted the number of occurrences for each word. The sentences had to elicit primarily the nouns they were associated with and only a small percentage (34%) of competitor or unrelated words.

The combination of these two norming studies was used to ensure that a given context was sufficiently associated with the target word, and not with the competitor word. Based on the norming data, we excluded eight pairs of words from the original collection: these were cases in which Strudel had produced properties for a different sense of the word than the one in the Huettig et al. materials, as well as cases in which the target sentences were too different from the competitor ones so that the properties could not be plausibly swapped.

The sentence materials were recorded by a native English speaker at a normal speech rate for presentation in the experiment.

7.1.1.2 Procedure

The entire experiment included 108 sentences: 18 word pairs (36 words in total) embedded in a neutral context and two biasing contexts. We rotated the position of the four objects on the screen to control for order or position effects. In total we therefore obtained 432 distinct items that we split in 24 lists of 18 items. The distribution of items across lists was based on a Latin square design, ensuring that each list included exactly one word from each target/competitor pair. Twenty-five filler items were added and a random presentation order generated for each list.

Twenty-four native English speakers from the University of Edinburgh were paid five pounds for taking part in the experiment. Each participant saw the items of one of the 24 lists, randomly interspersed with nine yes/no questions about the sentence or the scene. The questions were there to ensure that participants paid attention throughout the experiment.

The experiment setup is the same described in Section 5.3.1 for Experiment 2. The experiment lasted approximately 30 minutes.

7.1.2 Results

7.1.2.1 Fixation Probabilities

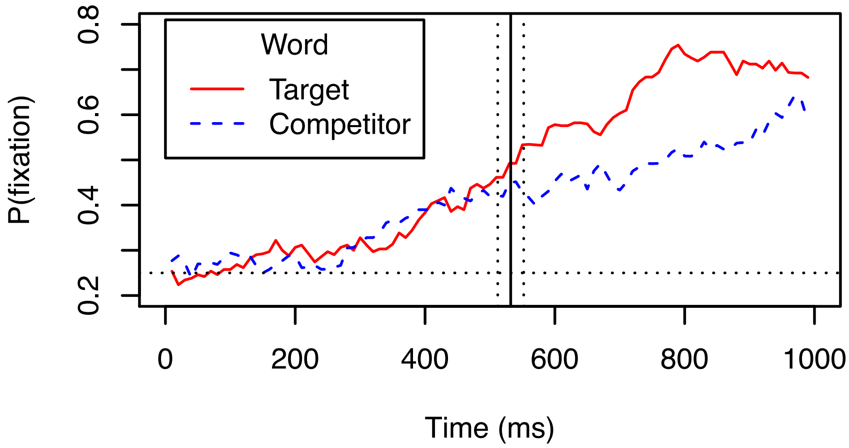
Our analysis is based on the fixations on the target object compared to the fixations on the three distractor objects on the display. As already anticipated, the contexts are less structured than the previous experiments. For this reason we can analyse the effects of context only at the time of the critical word when contextual information has already affected word's pre-activation. We excluded out-of-screen fixations and blinks from the analysis. Figure 7.2 plots the probability of fixating the target object across the three context conditions. The neutral context condition used the sentences of Huettig et al.; the target and competitor conditions used the contextually biased sentences produced based on the Strudel properties. In each plot, 0 ms corresponds to the acoustic onset of the critical word; our analysis takes into account the first 1000 ms after this onset. The vertical line shows the average offset of the critical words, with confidence intervals.

The horizontal line at .25 indicates the probability of randomly fixating one of the four objects. The alignment of the fixations in this experiment differs from the previous one where the fixations were aligned at the offset of the critical word.

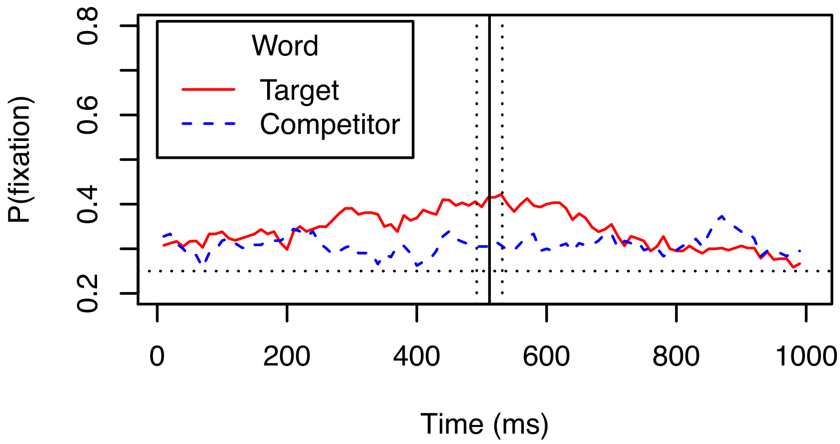
An inspection of the plots reveals a broadly similar trend across the three context conditions. The critical words require some time before they are recognised, which means that the fixation probabilities for the target and the competitor words take between 200 and 500 ms before they diverge. After that, we observe an increase in fixations to the target word compared to the competitor. The point of divergence is about 200 ms later in the neutral context; a semantically related context which is either the target or the competitor context seems to aid the recognition of the critical word and triggers early fixations to the corresponding object. (Bear in mind that the competitor context is also semantically related to the target, as our norming studies showed.)

Neutral Context Condition In the neutral context condition (Figure 7.2(a)), we observe a steady increase in fixation probability for both the target and the competitor word, which start to diverge at the offset of the critical word (this is presumably the point at which the critical word has been recognised by the participants). From that point on, we see more fixations on the target than on the competitor. This is in line with what Huettig et al. (2006) found: a competitor word triggers fixations to a semantically related target object, but less fixations than the target word corresponding to the target object. Our neutral context condition therefore provides a replication of Huettig et al.'s results. (The original paper also showed that the difference in fixation probability between target and competitor correlates with their semantic similarity, but we do not test this claim.)

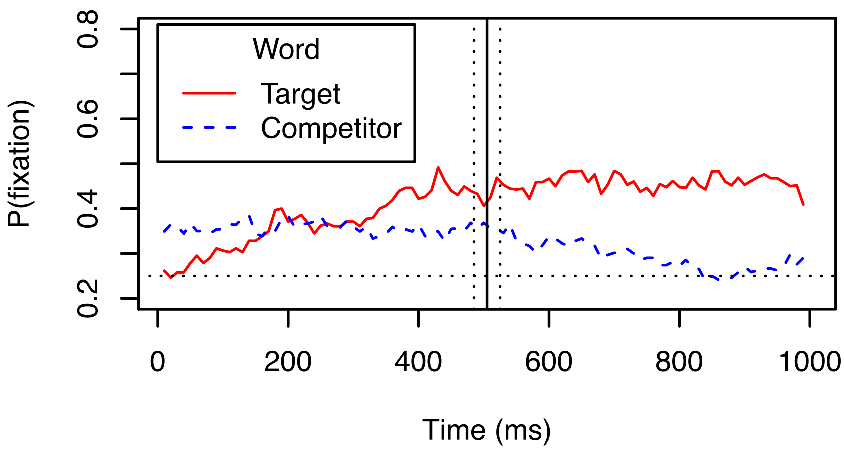
Target Context Condition In the target context condition (Figure 7.2(b)), participants hear a sentence containing properties of the depicted objects. Presumably this enables them to predict the target word with some accuracy (and our sentence completion study confirmed this). As the target is expected (and hence less interesting) at this point, we only observe a small increase of fixation probability for the target compared to the competitor, which starts early, at around 200 ms. This early start is consistent with the fact that participants are able to predict the critical word in this condition based on the context sentence.



(a) Fixation probability in a *neutral context* sentence.



(b) Fixation probability in a *target context* sentence.



(c) Fixation probability in *competitor context* sentence.

Figure 7.2: Fixation probabilities on the target object over time for the target (continuous red line) and competitor (dotted blue line) words. The onset of the critical word is at 0 ms. The vertical lines indicate the mean of the offset of the critical word with confidence interval. The horizontal line shows a probability of .25 (random baseline for four objects).

Competitor Context Condition In the competitor context condition (Figure 7.2(c)), participants hear a context sentence that is not directly associated with the depicted target object, but is instead associated with the semantically related competitor. In this case, hearing the target word (rather than the contextually appropriate competitor word) is unexpected, i.e., it generates interest and a larger increase in the number of fixations compared to the competitor word. This means that the two lines diverge more in the competitor context than in the target context condition, and the divergence remains high for the whole period of analysis.

7.1.2.2 Inferential Statistics

To statistically analyse the effect of the experimental manipulation on participants' fixations, we adopted the framework of linear mixed effect models (LME, Baayen, Davidson, & Bates, 2008).

The factor `Word` representing the nature of the critical word, coded as Competitor = $-.5$ and Target = .5. To determine context effects, we included two factors in helmert coding: the factor `Context` coded the difference between the neutral context = $-.5$ and the biasing context = .25 conditions; the factor `TargetSentence` differentiates the biasing context sentences further by distinguishing Competitor = $-.5$ and Target = .5. We have also included `Region` as a factor that indicates if the fixation is in the critical region (coded as $-.5$) or in the region after the offset of the critical word (coded .5). Finally, the continuous predictor `Time` was discretised into 10 ms bins (range 1–100).

We used the model selection procedure of Coco and Keller (2012) to find the minimal model that best fits our data. Table 7.1 gives the coefficients and significance levels for the minimal model; main effects or interactions not listed in this table were not included in the minimal model by the selection procedure. The random effects we included were `Participant` and `Item`, which were intercepts in the model. We also included random slopes for all the main effects (Word, Context, TargetSentence, Region, and Time).

Effect of Context The factor `Context` compares fixation probabilities in the neutral context and in the biasing context, collapsing the competitor and the target context in the biasing context condition. We find a significant, positive main effect of this factor, suggesting that participants make more fixations on the target object in the biasing context condition. This is modulated by a negative interaction `Time:Context`, which indicates that fixation probability increases over time in the neutral context condition.

Predictor	Coefficient
(Intercept)	-1.15^{***}
Time	0.17^{*}
Context	0.80^{*}
Time:Context	-0.64^{***}
TargetSentence	-0.47
Time:TargetSentence	0.11^{**}
Word	-0.06
Time:Word	0.18^{***}
Region	0.09
Region:Context	-0.41^{**}
Region:TargetSentence	-0.61^{***}
Word:TargetSentence	0.84^{***}
Time:Word:TargetSentence	-0.43^{***}
Region:Context:Word	-0.60^{***}

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 7.1: LME coefficients for the data in Figure 7.2.

This explains the upwards trend in Figure 7.2(a), but not in the biasing context conditions (Figures 7.2(b) and 7.2(c)).

While there is no general effect of whether the context is the competitor or the target sentence (no main effect of `TargetSentence`), we do find a significant positive interaction `Time:TargetSentence`. This confirms that there is a larger increase in fixations to the target object in the target context compared to the competitor context.

Effect of Critical Word While there is no main effect of `Word`, we find a significant positive interaction `Time:Word` that indicates that fixations on hearing the target word increase more quickly than fixations on hearing the competitor word. This is not surprising: when participants hear a word that matches the target object on the screen, they fixate this object more frequently (recall that the target object is depicted in all conditions, the competitor object is never on the screen).

Effect of Region There is no significant main effect of `Region`: whether the fixations are in the critical region (between the onset and the offset of the critical word) or in the post-critical region. However, we find a significant negative interaction `Region:Context`, suggesting that the neutral context sentences receive more fixations in the post-critical region compared to the biasing context sentences. This is compatible with the observation that context facilitates the processing of the critical word, which thus receives fixations earlier in the context condition.

The interaction `Region:TargetSentence` confirms that in the post-critical region participants fixate the target object more in the competitor context, presumably because it conflicts with their contextual expectations in this case. In the target context, however, contextual expectations and target object match, which means there is no reason to fixate the target object more frequently (compare Figures 7.2(b) and 7.2(c)).

Interaction of Context and Critical Word The most important interactions with respect to our experimental hypothesis are those involving `Context` and `Word` or `TargetSentence` and `Word`. These interactions demonstrate that context has an effect that is specific to the critical word.

We find a significant positive interaction `Word:TargetSentence`, which demonstrates that the target object receives more fixations when the target word occurs in the target context (rather than in the competitor context). This effect changes over time (significant negative interaction `Time:Word:TargetSentence`): the increase in fixations in the target word condition is larger in the competitor context than in the target context. For the competitor word, the opposite tendency emerges. This confirms the prediction that an expected critical word (i.e., one matching the context) is less interesting, and thus less likely to be fixated.

Finally, we can report a significant negative interaction `Region:Context:Word`, suggesting that the effect of `Word` in the neutral context condition is limited to the post-critical region, while in the biasing condition, it is stronger in the critical region. This corresponds to the observation that the fixation curves for the target and the competitor word diverge earlier for the biasing context conditions (see Figure 7.2).

7.1.3 Discussion

First of all, our results replicate the findings of Huettig et al. (2006). In the neutral context condition, we find that participants fixate the target object both when they hear the critical word, and when they hear the semantically related competitor. While we observe fewer fixations on the target for the competitor word, Figure 7.2(a) clearly indicates that it is fixated more frequently than chance (corresponding to a probability of .25).

However, the main purpose of our experiment was to analyse how the semantic similarity relations between context (neutral context, target context, competitor context) and target words affect the processing of the target word. We also test the cognitive validity of distributional semantic models when generating properties to use as experimental stimuli. We therefore included two context conditions in our experiment, one in which the context sentence contained properties related to the target word, and one in which it contained properties related to the competitor word. In this way we controlled for the semantic similarity between the entire context and the target/competitor word. In both cases, the properties were created by Strudel, a model of semantic representation.

When we compared these two biasing context conditions to the neutral context condition, we found two main effects. Firstly, a biasing context facilitates the processing of the critical word. Over time, the context builds up an expectation of the critical word, resulting in less fixations to the target object when it is contextually expected. This effect occurs for both types of biasing contexts, which is in line with the fact that the target and the competitor words were semantically related. In the neutral context, in contrast, no expectations can be computed, as participants cannot guess the identity of the target word before its onset. The target object is unexpected and hence more interesting and receives more fixations, but these fixations appear later, once the recognition of the target word is complete.

Our second finding is that a biasing context makes it possible to pre-activate the meaning of the critical word: in a target context, we get more fixations to the target during the target word, compared to the competitor word (Figure 7.2(b)). In the competitor context, we also initially find more fixations during the competitor word than during the target word. However, the pattern reverses after about 200 ms, presumably because of the match between the target word and the target object on the screen, which overrides the contextual expectation of the competitor word. Fixations for the target word remain high, however, compatible with a violation of contextual expectations (Figure 7.2(c)).

7.2 Experiment 6: Reducing the Influence of Visual Information

The comparison between the results obtained in the self-paced reading and the visual world paradigm experiments in Chapter 5, we showed a difference in the number of biasing words required to boost the activation of the target word. In the visual world paradigm, the visual information available for the duration of the linguistic stimulus provides a further source of facilitation and makes an early pre-activation of the target word possible. This property of the experimental paradigm provides an explanation for why participants are faster in pre-activating the target word in the eye-tracking study (after 1 HB context word) than in the reading study (after 2 HB context words).

A similar effect was also found in the results of the visual world paradigm experiment reported in the previous section: the pre-activation of the critical word driven by semantically similar contexts is influenced by the visual stimulus always depicted on the screen. The constant bottom-up support from the visual information is stronger for the target word than for the competitor word. For example, after a competitor context (see Figure 7.2(c)) the fixations are higher for the target word (matching with the visual stimulus) than for the competitor word (matching with the linguistic context).

Even though the presence of the visual stimulus is an intrinsic feature of the visual world paradigm, an established variant of the paradigm exists, in which the influence of the presented visual source is reduced: the blank screen paradigm. As described in Section 4.2, Altmann (2004) demonstrates that during language processing subjects direct their gaze towards an area of the screen that was previously occupied by an object related to the linguistic stimulus, even when the actual object is no longer depicted. The author claims that this effect is connected to the mental image that participants have constructed during the exposure to the visual stimulus.

In this section we replicate Experiment 5 using the blank screen paradigm. By removing the visual stimulus we encourage the process to move from a more visual modality to a more symbolic modality making stronger the top-down effect produced by the linguistic input. Considering that the persistence of the image in short term memory decreases over time (Vogel, Woodman, & Luck, 2001), the presence of a context that is related to the target object should maintain the mental image of the target object more vivid than the images of the other three distractors. On the contrary, a neutral sentence is expected to let the mental image of the visual scene decay over time.

The aim of this follow up experiment is to exclude, at least partially, the strong bias that the visual scene exerts on the interaction between context and target words.

7.2.1 Method

Materials and Procedure The materials, the procedure and the number of subjects are the same as in the previous experiment. The visual scene was on the screen for 5000 ms and, after that, it was substituted by a blank screen. At that point, the recorded sentence was played and followed by 1000 ms of silence before the end of the trial.

7.2.2 Results

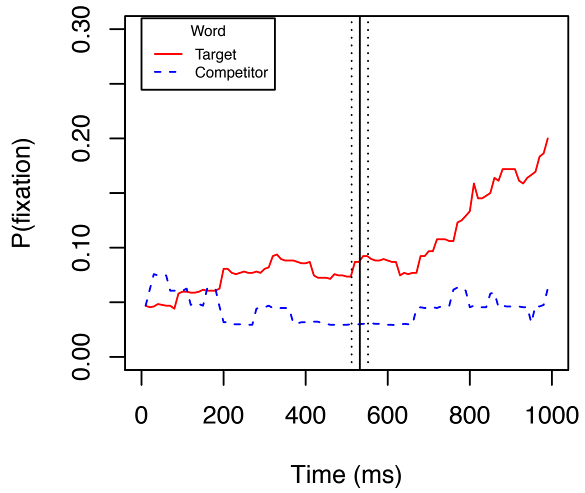
7.2.2.1 Fixation Probabilities

Figure 7.3 reports the probability of fixating the target object across the three context conditions. In each plot, 0 ms corresponds to the acoustic onset of the critical word; our analysis takes into account the first 1000 ms after this onset. The vertical line shows the average offset of the critical words, with confidence intervals. An analysis of the plots reveals a different trend across the three context conditions.

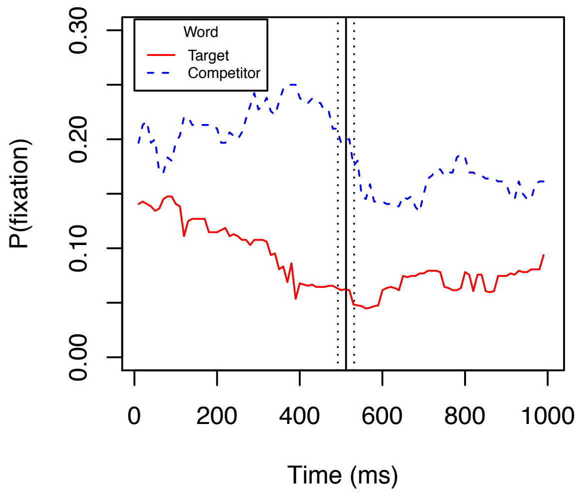
Neutral Context Condition In the neutral context condition (Figure 7.3(a)), we observe an increase in fixation probability for the target word and no effects for the competitor word. For the target word it is possible to identify an increase in fixations after 600 ms: probably the time when subjects have processed the critical word. No effect appears for the competitor condition. This is in line with the results of the previous experiment: even though later in time (600 ms here vs. 200 ms previously) a target word triggers fixations to the relevant target object. Unlike Experiment 5 (and Huettig et al.'s study), a competitor word does not trigger fixations to the target object. The relation between a competitor word (in a neutral context) and the mental image of the target object is not strong enough to affect word processing.

Target Context Condition Figure 7.3(b) shows that after a target context the probability of fixation decreases over time. The amount of fixations towards the target object is higher for the competitor word than for the target word; the fixations increase in the critical area and decrease after the offset.

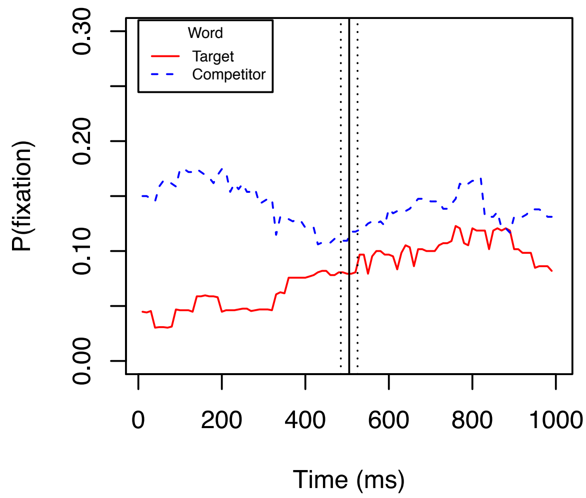
As discussed in the introduction of this study, the target context maintains the mental representation of the visual stimulus more vivid in memory. For this reason the



(a) Fixation probability in a *neutral context* sentence.



(b) Fixation probability in a *target context* sentence.



(c) Fixation probability in a *competitor context* sentence.

Figure 7.3: Fixation probabilities on the target object over time for the target (continuous red line) and competitor (dotted blue line) words. The onset of the critical word is at 0 ms. The vertical lines indicate the mean of the offset of the critical word with confidence interval.

target context has already activated the processing of the target word. The absence of the visual stimulus does not bias participants to fixate again the area where the target object was depicted because the target object has been already identified during the processing of contextual information. On the other hand, the expectations generated by the target context do not correspond to the competitor word. This causes an increase in the number of fixations because participants are evaluating the mismatch between context and critical word.

Competitor Context Condition In the competitor context condition (Figure 7.3(c)), the trend of the two curves is similar to the trend described in the previous experiment: an increase of fixations during the production of the target word, and no effects (with a slight negative slope) for the competitor word. The target word is unexpected and it generates interest and a steeper slope in the number of fixations compared to the competitor word that is contextually expected. It is interesting to note that even though the trend is the same as the one in the previous experiment, the position of the curves is reversed: after a competitor context the competitor word produces more fixations than the target one (even though this difference is not significant). We can explain this difference based on the weaker effect of the visual stimulus. In the previous experiment, we showed that after a competitor context the amount of fixations is higher for the competitor word than for the target word; this trend reverses after 200 ms because of the strong effect of the visual stimulus. The absence of the visual stimulus in this experiment does not produce the same result.

7.2.2.2 Inferential Statistics

In this section we report the LME analysis for the collected data. The dependent variable was the empirical logit of the fixation probability. The fixed and random factors of this model were the same as described in section 7.1.2.2.

Table 7.2 reports the coefficients and significance levels for the minimal model; main effects or interactions not listed in this table were not included in the minimal model by the selection procedure.

Effect of Context The factor `Context` compares fixation probabilities in the neutral context and in the biasing context (target and competitor together). The significant, positive main effect of this factor suggests that participants fixate more often the target object in the biasing contexts than in the neutral condition. The negative interaction

Predictor	Coefficient
(Intercept)	-2.38^{***}
Time	-0.01
Context	1.11^{***}
Time:Context	-0.28^{***}
TargetSentence	0.12
Word	-0.43^{*}
Time:Word	0.10^{***}
Region	0.02
Region:Context	0.50^{***}
Region:TargetSentence	-0.47^{***}
Context:Word	-1.12^{***}
Word:TargetSentence	0.51^{***}
Time:Word:TargetSentence	-0.27^{***}
Region:Context:Word	-0.41^{***}
Region:TargetSentence:Word	0.42^{**}

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 7.2: LME coefficients for the data in Figure 7.3.

`Time:Context` indicates that fixation probability increases over time in the neutral context condition. This reflects the upwards trend in Figure 7.3(a), but not in the biasing context conditions (Figures 7.3(b) and 7.3(c)). These outcomes correspond to the results of the analysis reported in the previous experiment. There is no general effect of whether the context is the competitor or the target sentence (no main effect of `TargetSentence`).

Effect of Critical Word There is a significant negative main effect of `Word` indicating more fixations for the competitor word than the target word; and a significant positive interaction `Time:Word` that indicates that fixations on the target word increase more than fixations on the competitor word (similar to the previous experiment). When participants hear a word that matches with one of the entities in the mental image of the visual scene, they fixate that area of the screen more frequently. There is a negative interaction `Context:Word` that shows more fixations during the target word after a neutral context than after a biasing context.

Effect of Region There is no significant main effect of `Region`. However, we find a significant positive interaction `Region:Context`, suggesting that the biasing context sentences receive more fixations in the post-critical region compared to the neutral context sentences. This interaction have an opposite sign in the previous study. This effect is related to the very low amount of fixations for the competitor word in the neutral context. The interaction `Region:TargetSentence` confirms that in the post-critical region participants fixate the target object more in the competitor context than in the target context, presumably because it conflicts with the expectations generated by the context. While in the target context, contextual expectations and the target object correspond; this elicits a smaller number of fixations.

Interaction of Context and Critical Word The `Context` and `Word` interaction and the `TargetSentence` and `Word` interaction demonstrate that context has an effect that is specific to the critical word. We find a significant positive interaction `Word:TargetSentence`, which demonstrates that the target object receives more fixations when the target word occurs in the target context (rather than in the competitor context). This effect changes over time (significant negative interaction `Time:Word:TargetSentence`): the increase in fixations in the target word condition is larger in the competitor context than in the target context. For the competitor word, the opposite tendency emerges. This confirms the prediction that an expected critical word (i.e., one matching the context) is less interesting, and thus less likely to be fixated. We report a significant negative interaction `Region:Context:Word`, suggesting that the effect of `Word` in the neutral context condition is limited to the post-critical region, while in the biasing condition, it is stronger in the critical region. This corresponds to the observation that the fixation curves for the target and the competitor word do never overlap in biasing context conditions. Finally, the positive interaction `Region:TargetSentence:Word` shows that there is a stronger difference between the fixations associated with target and competitor words in the post-critical region of the target context.

7.2.3 Discussion

In this section we report the results of an experiment that used the blank screen paradigm to mitigate the effect of the visual scene on the interaction of context and target words.

The absence of the visual stimulus produces a significant reduction in the fixations in the direction of the target object. The analysis of the raw data shows a preference in fixating the centre, the top or the bottom of the screen when the picture is not depicted.

In the introduction we suggested that a biasing context keeps the mental image of the visual scene more vivid than a neutral context. This is supported by the results obtained in this study. The amount of fixations in the neutral context is lower than the amount of fixations in the biasing contexts. Moreover, the very low number of fixations for the competitor word suggests that the semantic similarity between the competitor word and the mental image of the visual scene is not strong enough to produce fixations towards the area of the screen where the target object was depicted. This last result does not replicate the outcomes of Experiment 5 and of the original experiment performed by Huettig et al. (2006) where also a competitor word elicits a high amount of fixations towards the semantic similar object depicted on the screen. Moreover, the target context makes the mental image (an the consequent spatial position of the target object on the screen) more vivid in mind. This allows the pre-activation of the target word. When hearing the target word participants are no more looking at the position where the target object was depicted given that (compared to the previous study) there is not a visual stimulus available. While, when hearing a competitor word that contradicts the expectations, the amount of fixations is higher. Finally, the competitor context reduces the influence of the visual image making the linguistic relation occurring between the context and the target word stronger.

The results obtained in the blank screen paradigm experiment confirm the validity of our claim: the existence of an expectation produced by contextual information can be observed also when the visual scene is not present. Not surprisingly, the combination of linguistic and visual information generates a stronger effect: the identification of the target object takes place earlier in time. Moreover, the presence of the visual stimulus attracts more fixations also after the activation of the critical word.

Overall, the results of this experiment only partially replicate the outcomes of Experiment 5 because the amount of fixations is very low and the data in general more noisy.

Even if the results of this experiment do not allow us to make any strong claim about the effect of visual information on language processing we believe that they represent a worthy methodological contribution to the visual world experimental paradigm.

7.3 Conclusion

The effect of context on semantically similar words was the focus of this chapter. The properties (context words) used to test this hypothesis were generated by Strudel, a distributional semantic model that represents word meanings in terms of feature norm-like properties. With this model we could generate properties also for words that were not part of existing collections of norms. The produced contexts are semantically similar to the critical word they refer to. Moreover, we used semantically similar pairs of words as critical words under the assumption that those words share a certain amount of properties that the context can pre-activate.

In Experiment 4 we analysed the fixations towards the target object at the time of the critical word. This word was following a neutral context, a related context, or a context related to a semantic similar word. Overall, the results of this experiment have confirmed the outcomes of Federmeier and Kutas (1999) and Huettig et al. (2006): words that are semantically related produce similar effects in eye-movements. Both target and competitor critical words produced an increasing amount of fixations towards the target object depicted on the screen. This evidence is in favour of a pre-activation effect driven by context at the property level. The trend of the fixations described in this experiment is similar to the trend described in Experiment 2 when analysing the fixations at the critical word: stronger is the relation between context and target words lower is the amount of fixations generated.

In Experiment 5 we partially replicate the general outcomes of Experiment 4 in a blank screen paradigm experiment. The aim of this experiment was to reduce the strong effect of the visual information while the context was unfolding. According to this paradigm, participants generate a mental representation of the visual information provided and use this representation to direct their eye-movements while the visual scene is removed. The lack of visual information causes an overall reduction in the total number of fixations. We show that a biasing context makes the visual scene more vivid in participant's minds. This effect is described by the higher amount of fixations after a biasing context.

Experiments 4 and 5 agree with the predictive account. A coherent context allows the pre-activation (earlier divergence point in Experiment 5) of the critical word and consequently increases the expectation (less fixations in both experiments) of it. Compared to Experiment 2, we cannot directly analyse the time course when context was produced. For this reason we cannot produce any evidence in favour of the immediate or the integrative activation hypotheses discussed in Chapter 5.

Finally, both experiments confirm the claim that distributional models of semantics can generate properties that are cognitively plausible. They are plausible in the sense that they can be used to construct contexts that successfully bias participants towards an object that is compatible with the context.

Chapter 8

Conclusion

In this thesis we investigated how word processing is affected by incremental contextual information. The main assumption of the predictive account (see Chapter 2) is that context produces a facilitation effect that is additive and occurs together with the unfolding sentence. The aim of this work was to test the predictive account assumption by manipulating the number of biasing words in the context.

We formulated two hypotheses that describe the relation between the amount of biasing context provided and the pre-activation generated. The *incremental activation hypothesis* completely follows the assumptions of the predictive account: the context has an instantaneous effect on word processing and this effect is additive. On the other hand, *the immediate activation hypothesis* suggests that a certain amount of biasing information is required to boost the activation of the critical word. After the expected word has been pre-activated, increasing the amount of context does not produce any additional significant facilitation effect. We tested the different effect of context predicted by the two hypotheses also in terms of feature overlap: as the context unfolds, the semantic features of the processed words are activated and the upcoming words that match those features are pre-activated and thus processed more quickly when encountered.

8.1 Contributions

In order to test the main assumption of the predictive account, we performed six experiments where we analysed the facilitation effect of context over time manipulating the amount of biasing context provided. These studies allowed us to test the predictive account in terms of the two activation hypotheses. In a self-paced reading study

(**Experiment 1**) we showed that at least two biasing words (a location and an actor) are required in order to facilitate the processing of the critical word. In a visual world study (**Experiment 2**) we investigated the facilitation effect as the context was unfolding. The results showed that the first biasing context word (a location) already pre-activated the meaning of the critical word. The inclusion of more contextual information did not produce any additional significant effect. The different number of biasing words required to boost the activation shown in the two experiments is directly related to the different methodology used to perform the experiments. The constant presence of a target visual stimulus reduces the amount of linguistic information required to produce facilitation. Taken together, the results of these two experiments agree with the predictive account showing that the pre-activation of the critical word is driven by contextual information. According to the immediate activation hypothesis the effect of context does not immediately affect the pre-activation of word meaning and this effect is not additive.

In a second self-paced reading study (**Experiment 3**) we investigated how the order of the words in the context affects the time required to process the target word. The significant differences between conditions that emerged in the previous studies were not replicated. We identified the different syntactic structure of the linguistic stimuli as the reason for this outcome.

Finally, in four association studies (**Experiment 4**) we analysed the semantic relations between each context word and the critical one, and between each context word independently. These association studies showed that the context words that are highly associated to the target word are also strongly associated to each other. The outcome of these studies provides a possible explanation to the effect of context described in Experiment 1 and Experiment 2 in terms of the immediate activation hypothesis. The strong relation among the HB context words restricts the number of expected candidates in the continuation of the sentence and facilitates the immediate pre-activation of the right candidate.

In Chapter 6 we used a bag-of-words distributional semantic model to address the feature overlap theory. Each dimension of the vectors generated by the model can be treated as a semantic property. A series of correlation studies showed that the semantic similarity scores between the word vectors computed by the model significantly correlate with the association scores produced by humans. We demonstrated that it is possible to model the feature overlap theory in terms of vector combination. When we used the point-wise multiplication to combine the vectors of each context word, the model successfully predicted the reading times of Experiment 1 by assuming that words with higher similarity with the context would be read faster.

Finally, in Chapter 7 we performed two new eye-tracking studies to analyse the contextual effects on semantically similar words. In a visual world study (**Experiment 5**) we showed that semantically similar words produce similar patterns in eye-movements. A coherent context allows the pre-activation of the critical word and it increases the expectations towards it. We show that the fixations are directly connected with the semantic similarity between the context words and the critical word. We performed a blank screen paradigm study (**Experiment 6**) in order to reduce the strong facilitation effect of the visual information on the linguistic task. This study only partially replicates the results of Experiment 5; it shows that the facilitation effect is driven by the linguistic context even when the visual information is not directly available on the screen. In these two experiments we also tested whether a distributional model could generate words that can be successfully embedded into experimental stimuli, and found that these words bias fixations towards a related object depicted on the screen (demonstrating the cognitive validity of the model).

8.2 Future work

The experiments described in this thesis have analysed the interaction between context and word meaning; however there are open questions.

8.2.1 Manipulating the Order of Context Words

In Section 5.4 we reported a self-paced reading experiment where we manipulated the order of the three context words (location, actor, object). The study was meant to investigate the effect of the order in which context words are presented. We aimed to understand if the results obtained in Experiment 1 and Experiment 2 were due to the semantic nature of the first word provided (a location) or by the fact that the word was the first high biasing word provided. In order to construct acceptable English sentences, we changed the syntax of the linguistic stimuli by introducing a subordinate clause that embedded the critical word. The facilitation effect driven by a biasing context that we showed in the previous studies was not replicated in this study. We discussed the different syntactic structure of the sentence as the most possible reason for these results. A possible way to solve this problem is to construct linguistic stimuli that include context and target words in the main clause. It would be still possible to design the linguistic stimuli in English, however the fixed word order characterising this language would probably affect the quality of the final materials. A solution can be the use of lists of words instead of sentences as stimuli in a reading task. Otherwise, the use of a language that imposes less syntactic constraints, for example German, would guarantee more flexibility in word order and would allow a more free manipulation of the context words. Based on the outcomes of the experiments described in this thesis, we expect to see that the order and the semantic type of the words do not significantly change the effect of context.

8.2.2 Constructing Contexts with Unrelated Biasing Words

In the association studies reported in Section 5.5.2 we showed that words that are highly related to the critical word are also highly related to each other. This strong inter-contextual relation can explain the ceiling effect that emerged in Experiment 2: when the critical word has been pre-activated, increasing the amount of biasing information do not cause any additional significant facilitation effect. We explained this effect in terms of feature overlap. The overlap between strongly related HB words is very high and causes an immediate activation of the target word. In order to manipulate the relation between context words a new set of linguistic stimuli should include words that are highly related to the critical word but not strongly related to each other. The best candidates to insert into the linguistic stimuli can be identified using the similarity scores computed by a distributional semantic model. As we saw in Chapter 6, the similarity scores correlate with the association judgements produced by humans. When using context words that are weakly related to one another we expect to see a pattern more coherent with the incremental activation hypothesis described in Chapter 5 where the effect of context is additive.

8.2.3 Analysing Incrementality in the Visual Scene

So far we have investigated the effect of linguistic context on the activation and processing of word meaning. However, when interacting with real world situations, a strong facilitation effect is driven by the visual information (Oliva & Torralba, 2007). In the future we would like to investigate whether visual context is processed in similar ways as the thesis has shown linguistic context to be used. As the results of the blank screen paradigm experiment suggested, the inclusion of biasing information in the visual scene should produce stronger facilitation effects than those produced by the linguistic context alone. A possible direction for this study would be to manipulate incrementality both in the visual and in the linguistic context. The linguistic stimuli would be the same as those used in the previous experiments. The visual stimuli should be constructed in order to contain objects strongly or weakly related to the linguistic stimuli. This experiment would allow us to analyse the interaction of visual and linguistic information in facilitating word processing.

Appendix A

Experimental Materials

A.1 Linguistic Stimuli used in Experiments 1 and 2

We provide the linguistic stimuli used in Experiment 1 and Experiment 2 (see Chapter 5). We report only the conditions with three HB context words (a) and three LB context words (b). The different context words are marked with their semantic type ("@loc" indicates the location, "@act" indicates the actor, and "@obj" indicates the object). The target word is marked with the string "@target". In the self-paced reading experiment we had 8 conditions in total (see Section 5.2) where we produced all the possible combinations of HB and LB context words. In the visual world paradigm experiment we reduced the number of conditions to 4 (see Section 5.3).

1. (a) Before the folk_concert@loc the musician@act was checking the keys@obj of the accordion@target for a couple of minutes.

   (b) Before the party@loc the man@act was checking the case@obj of the accordion@target for a couple of minutes.

2. (a) At the stop@loc the driver@act was picking up some passengers@obj on the bus@target quickly.

   (b) On the road@loc the man@act was picking up some people@obj on the bus@target quickly.

3. (a) In the fortress@loc the soldier@act was inserting the ammunition@obj into the cannon@target carefully.

   (b) In the building@loc the man@act was inserting the powder@obj into the cannon@target carefully.

4. (a) In the parking_lot@loc the driver@act was wiping the rims@obj of the car@target carefully.
   (b) In the square@loc the man@act was wiping the windows@obj of the car@target carefully.

5. (a) In the kitchen@loc the chef@act was cutting some sticks@obj of carrot@target carefully.
   (b) In the room@loc the man@act was cutting some pieces@obj of carrot@target carefully.

6. (a) In the park@loc the gardener@act was removing some larvae@obj of caterpillar@target carefully.
   (b) In the field@loc the man@act was removing some eggs@obj of caterpillar@target carefully.

7. (a) For the reception@loc the gentleman@act was buying the hat@obj and the coat@target half price.
   (b) For the dinner@loc the mechanic@act was buying the perfume@obj and the coat@target half price.

8. (a) In the kitchen@loc the chef@act was using the pan@obj on the cooker@target carefully.
   (b) In the room@loc the man@act was using the spoon@obj on the cooker@target carefully.

9. (a) At the mill@loc the farmer@act was separating the straw@obj from the corn@target for later.
   (b) In the field@loc the man@act was separating the grass@obj from the corn@target for later.

10. (a) Before the parade@loc the musician@act was preparing the sticks@obj and the drum@target for later.
    (b) On the road@loc the man@act was preparing the string@obj and the drum@target for later.

11. (a) In the jungle@loc the poacher@act was carrying the tusks@obj of the elephant@target for a long time.

(b) In the grass@loc the man@act was carrying the teeth@obj of the elephant@target for a long time.

12. (a) In the pub@loc the barman@act was taking some bottles@obj out of the fridge@target quickly.

    (b) In the room@loc the man@act was taking some water@obj out of the fridge@target quickly.

13. (a) In the studio@loc the player@act was fixing the strings@obj of the guitar@target for later.

    (b) In the room@loc the man@act was fixing the case@obj of the guitar@target for later.

14. (a) Before the concert@loc the musician@act was replacing a string@obj of the harp@target for later.

    (b) In the room@loc the man@act was replacing a screw@obj of the harp@target for later.

15. (a) In the cafeteria@loc the waiter@act was putting some tea@obj in the kettle@target for later.

    (b) In the room@loc the man@act was putting some water@obj in the kettle@target for later.

16. (a) On the ice_rink@loc the skater@act was holding the scarf@obj and the mittens@target for a while.

    (b) On the field@loc the man@act was holding the jacket@obj and the mittens@target for a while.

17. (a) In the forest@loc the picker@act was holding a basket@obj full of mushrooms@target carefully.

    (b) On the path@loc the man@act was holding a box@obj full of mushrooms@target carefully.

18. (a) In the zoo@loc the keeper@act was looking at the feathers@obj of the peacock@target for a while.

    (b) In the field@loc the man@act was looking at the neck@obj of the peacock@target for a while.

19. (a) At the concert@loc the musician@act was checking the keys@obj of the piano@target carefully.
    (b) In the room@loc the man@act was checking the cover@obj of the piano@target carefully.
20. (a) In the shop@loc the chef@act was buying the peeler@obj and the potatoes@target for later.
    (b) In the building@loc the man@act was buying the knife@obj and the potatoes@target for later.
21. (a) In the garden@loc the gardener@act was looking for the shell@obj of the snail@target for a while.
    (b) On the path@loc the kid@act was looking for the trail@obj of the snail@target for a while.
22. (a) In the kitchen@loc the cook@act was taking some bread@obj out of the toaster@target quickly.
    (b) In the room@loc the man@act was taking some crumbs@obj out of the toaster@target quickly.
23. (a) At the station@loc the conductor@act was inspecting the hitches@obj of the train@target carefully.
    (b) At the stop@loc the man@act was inspecting the lights@obj of the train@target carefully.
24. (a) At the stadium@loc the referee@act was holding the yellow_card@obj and the whistle@target for a while.
    (b) In the park@loc the man@act was holding the watch@obj and the whistle@target for a while.

A.2 Linguistic Stimuli used in Experiment 3

We provide the linguistic stimuli used in Experiment 3 (see Chapter 5). We report only the conditions with three HB context words (a-c-e) and three LB context words (b-d-f) when the location is the first context word (a-b), when the actor is the first context word (c-d), and when the object is the first context word (e-f). The different context words are marked with their semantic type ("@loc" indicates the location, "@act" indicates the actor, and "@obj" indicates the object). The target word is marked with the string "@target". In the reading experiment other possible combinations of HB and LB context words were also included (HB-LB-LB, HB-HB-LB).

1. (a) Before the folk_concert@loc the musician@act was checking the keys@obj, while the accordion@target was being cleaned.
   (b) Before the party@loc the man@act was checking the case@obj, while the accordion@target was being cleaned.
   (c) The musician@act before the folk_concert@loc was checking the keys@obj, while the accordion@target was being cleaned.
   (d) The man@act before the party@loc was checking the case@obj, while the accordion@target was being cleaned.
   (e) The keys@obj were being checked by the musician@act before the folk_concert@loc, while the accordion@target was cleaned.
   (f) The case@obj was being checked by the man@act before the party@loc, while the accordion@target was cleaned.
2. (a) At the stop@loc the driver@act picked up some passengers@obj, since the bus@target was not yet full.
   (b) On the road@loc the man@act picked up some people@obj, since the bus@target was not yet full.
   (c) The driver@act at the stop@loc picked up some passengers@obj, since the bus@target was not yet full.
   (d) The man@act on the road@loc picked up some people@obj, since the bus@target was not yet full.
   (e) Some passengers@obj were being picked up by the driver@act at the stop@loc, since the bus@target was not yet full.
   (f) Some people@obj were being picked up by the man@act on the road@loc, since the bus@target was not yet full.
3. (a) In the fortress@loc the soldier@act loaded the ammunition@obj, while the cannon@target was being readied carefully.

(b) In the building@loc the man@act loaded the powder@obj, while the cannon@target was being readied carefully.

(c) The soldier@act in the fortress@loc loaded the ammunition@obj, while the cannon@target was being readied carefully.

(d) The man@act in the building@loc loaded the powder@obj, while the cannon@target was being readied carefully.

(e) The ammunition@obj was being loaded by the soldier@act in the fortress@loc, while the cannon@target was readied carefully.

(f) The powder@obj was being loaded by the man@act in the building@loc, while the cannon@target was readied carefully.

4. (a) In the parking_lot@loc the driver@act wiped the rims@obj, while the car@target was sitting with its engine off.

(b) In the square@loc the man@act wiped the windows@obj, while the car@target was sitting with its engine off.

(c) The driver@act in the parking_lot@loc wiped the rims@obj, while the car@target was sitting with its engine off.

(d) The man@act in the square@loc wiped the windows@obj, while the car@target was sitting with its engine off.

(e) The rims@obj were being wiped by the driver@act in the parking_lot@loc, while the car@target was sitting with its engine off.

(f) The windows@obj were being wiped by the man@act in the square@loc, while the car@target was sitting with its engine off.

5. (a) In the kitchen@loc the chef@act was cutting some sticks@obj, while the carrot@target was on the board.

(b) In the room@loc the man@act was cutting some pieces@obj, while the carrot@target was on the board.

(c) The chef@act in the kitchen@loc was cutting some sticks@obj, while the carrot@target was on the board.

(d) The man@act in the room@loc was cutting some pieces@obj, while the carrot@target was on the board.

(e) Some sticks@obj were cut by the chef@act in the kitchen@loc, while the carrot@target was on the board.

(f) Some pieces@obj were cut by the man@act in the room@loc, while the carrot@target was on the board.

6. (a) In the park@loc the gardener@act was removing some larvae@obj, while the caterpillar@target was crawling away.

(b) In the field@loc the man@act was removing some eggs@obj, while the caterpillar@target was crawling away.

(c) The gardener@act in the park@loc was removing some larvae@obj, while the caterpillar@target was crawling away.

(d) The man@act in the field@loc was removing some eggs@obj, while the caterpillar@target was crawling away.

(e) Some larvae@obj were being removed by the gardener@act from the park@loc, while the caterpillar@target was crawling away.

(f) Some eggs@obj were being removed by the man@act from the field@loc, while the caterpillar@target was crawling away.

7. (a) For the reception@loc the gentleman@act bought the hat@obj, while the coat@target was being cleaned.

(b) For the dinner@loc the mechanic@act bought the perfume@obj, while the coat@target was being cleaned.

(c) The gentleman@act for the reception@loc bought the hat@obj, while the coat@target was being cleaned.

(d) The mechanic@act for the dinner@loc bought the perfume@obj, while the coat@target was being cleaned.

(e) The hat@obj was being bought by the gentleman@act for the reception@loc, while the coat@target was cleaned.

(f) The perfume@obj was being bought by the mechanic@act for the dinner@loc, while the coat@target was cleaned.

8. (a) In the kitchen@loc the chef@act was preparing the pan@obj, while the cooker@target was heating up.

(b) In the room@loc the man@act was preparing the spoon@obj, while the cooker@target was heating up.

(c) The chef@act in the kitchen@loc was preparing the pan@obj, while the cooker@target was heating up.

(d) The man@act in the room@loc was preparing the spoon@obj, while the cooker@target was heating up.

(e) The pan@obj was prepared by the chef@act in the kitchen@loc, while the cooker@target was heating up.

(f) The spoon@obj was prepared by the man@act in the room@loc, while the cooker@target was heating up.

9. (a) At the mill@loc the farmer@act was gathering the straw@obj, while the corn@target was being stored for later.

   (b) In the field@loc the man@act was gathering the grass@obj, while the corn@target was being stored for later.

   (c) The farmer@act at the mill@loc was gathering the straw@obj, while the corn@target was being stored for later.

   (d) The man@act in the field@loc was gathering the grass@obj, while the corn@target was being stored for later.

   (e) The straw@obj was being gathered by the farmer@act at the mill@loc, while the corn@target was stored for later.

   (f) The grass@obj was being gathered by the man@act in the field@loc, while the corn@target was stored for later.

10. (a) Before the parade@loc the musician@act prepared the sticks@obj, while the drum@target sat nearby.

    (b) On the road@loc the man@act prepared the string@obj, while the drum@target sat nearby.

    (c) The musician@act before the parade@loc prepared the sticks@obj, while the drum@target sat nearby.

    (d) The man@act on the road@loc prepared the string@obj, while the drum@target sat nearby.

(e) The sticks@obj were being prepared before the parade@loc by the musician@act, while the drum@target sat nearby.

(f) The string@obj was being prepared on the road@loc by the man@act, while the drum@target sat nearby.

11. (a) In the jungle@loc the poacher@act was carrying the tusks@obj, while the elephant@target was running away.

(b) In the grass@loc the man@act was carrying the teeth@obj, while the elephant@target was running away.

(c) The poacher@act in the jungle@loc was carrying the tusks@obj, while the elephant@target was running away.

(d) The man@act in the grass@loc was carrying the teeth@obj, while the elephant@target was running away.

(e) The tusks@obj were being carried by the poacher@act in the jungle@loc, while the elephant@target was running away.

(f) The teeth@obj were being carried by the man@act in the grass@loc, while the elephant@target was running away.

12. (a) In the pub@loc the barman@act took out some bottles@obj, while the fridge@target was open.

(b) In the room@loc the man@act took out some water@obj, while the fridge@target was open.

(c) The barman@act in the pub@loc took out some bottles@obj, while the fridge@target was open.

(d) The man@act in the room@loc took out some water@obj, while the fridge@target was open.

(e) Some bottles@obj were being taken out by the barman@act in the pub@loc, while the fridge@target was open.

(f) Some water@obj was being taken out by the man@act in the room@loc, while the fridge@target was open.

13. (a) In the studio@loc the player@act was fixing the strings@obj, while the guitar@target was laying on the table.

(b) In the room@loc the man@act was fixing the case@obj, while the guitar@target was laying on the table.

(c) The player@act in the studio@loc was fixing the strings@obj, while the guitar@target was laying on the table.

(d) The man@act in the room@loc was fixing the case@obj, while the guitar@target was laying on the table.

(e) The strings@obj were being fixed by the player@act in the studio@loc, while the guitar@target was laying on the table.

(f) The case@obj was being fixed by the man@act in the room@loc, while the guitar@target was laying on the table.

14. (a) Before the concert@loc the musician@act replaced a string@obj, while the harp@target was being polished.

(b) In the room@loc the man@act replaced a screw@obj, while the harp@target was being polished.

(c) The musician@act before the concert@loc replaced a string@obj, while the harp@target was being polished.

(d) The man@act in the room@loc replaced a screw@obj, while the harp@target was being polished.

(e) A string@obj was being replaced by the musician@act before the concert@loc, while the harp@target was being polished.

(f) A screw@obj was being replaced by the man@act in the room@loc, while the harp@target was being polished.

15. (a) In the cafeteria@loc the waiter@act was serving some tea@obj, while the kettle@target was cooling down.

(b) In the room@loc the man@act was serving some water@obj, while the kettle@target was cooling down.

(c) The waiter@act in the cafeteria@loc was serving some tea@obj, while the kettle@target was cooling down.

(d) The man@act in the room@loc was serving some water@obj, while the kettle@target was cooling down.

(e) Some tea@obj was being served by the waiter@act in the cafeteria@loc, while the kettle@target was cooling down.

(f) Some water@obj was being served by the man@act in the room@loc, while the kettle@target was cooling down.

16. (a) On the ice_rink@loc the skater@act was wearing the scarf@obj, while the mittens@target were lying on the ice.

(b) On the field@loc the man@act was wearing the jacket@obj, while the mittens@target were lying on the ice.

(c) The skater@act on the ice_rink@loc was wearing the scarf@obj, while the mittens@target were lying on the ice.

(d) The man@act on the field@loc was wearing the jacket@obj, while the mittens@target were lying on the ice.

(e) The scarf@obj was held by the skater@act on the ice_rink@loc, while the mittens@target were lying on the ice.

(f) The jacket@obj was held by the man@act on the field@loc, while the mittens@target were lying on the ice.

17. (a) In the forest@loc the picker@act was holding a basket@obj, while the mushrooms@target were being picked carefully.

(b) On the path@loc the man@act was holding a box@obj, while the mushrooms@target were being picked carefully.

(c) The picker@act in the forest@loc was holding a basket@obj, while the mushrooms@target were being picked carefully.

(d) The man@act on the path@loc was holding a box@obj, while the mushrooms@target were being picked carefully.

(e) A basket@obj was held by the picker@act in the forest@loc, while the mushrooms@target were being picked carefully.

(f) A box@obj was held by the man@act on the path@loc, while the mushrooms@target were being picked carefully.

18. (a) In the zoo@loc the keeper@act was admiring the feathers@obj, while the peacock@target was roosting quietly.

(b) In the field@loc the man@act was admiring the neck@obj, while the peacock@target was roosting quietly.

(c) The keeper@act in the zoo@loc was admiring the feathers@obj, while the peacock@target was roosting quietly.

(d) The man@act in the field@loc was admiring the neck@obj, while the peacock@target was roosting quietly.

(e) The feathers@obj were being admired by the keeper@act in the zoo@loc, while the peacock@target was roosting quietly.

(f) The neck@obj was being admired by the man@act in the field@loc, while the peacock@target was roosting quietly.

19. (a) At the concert@loc the musician@act was checking the keys@obj, while the piano@target was being polished for the event.

(b) In the room@loc the man@act was checking the cover@obj, while the piano@target was being polished for the event.

(c) The musician@act at the concert@loc was checking the keys@obj, while the piano@target was being polished for the event.

(d) The man@act in the room@loc was checking the cover@obj, while the piano@target was being polished for the event.

(e) The keys@obj were checked by the musician@act at the concert@loc, while the piano@target was being polished for the event.

(f) The cover@obj was checked by the man@act in the room@loc, while the piano@target was being polished for the event.

20. (a) In the shop@loc the chef@act bought the peeler@obj, while the potatoes@target were being washed at the restaurant.

(b) In the building@loc the man@act bought the knife@obj, while the potatoes@target were being washed at the restaurant.

(c) The chef@act in the shop@loc bought the peeler@obj, while the potatoes@target were being washed at the restaurant.

(d) The man@act in the building@loc bought the knife@obj, while the potatoes@target were being washed at the restaurant.

(e) The peeler@obj was being bought by the chef@act in the shop@loc, while the potatoes@target were washed at the restaurant.

(f) The knife@obj was being bought by the man@act in the building@loc, while the potatoes@target were washed at the restaurant.

21. (a) In the garden@loc the gardener@act was searching for the shell@obj, while the snail@target was finding food.

(b) On the path@loc the kid@act was searching for the trail@obj, while the snail@target was finding food.

(c) The gardener@act in the garden@loc was searching for the shell@obj, while the snail@target was finding food.

(d) The kid@act on the path@loc was searching for the trail@obj, while the snail@target was finding food.

(e) The shell@obj was being sought by the gardener@act in the garden@loc, while the snail@target was finding food.

(f) The trail@obj was being sought by the kid@act on the path@loc, while the snail@target was finding food.

22. (a) In the kitchen@loc the cook@act took the bread@obj out quickly, as the toaster@target was giving off a burnt smell.

(b) In the room@loc the man@act took the crumbs@obj out quickly, as the toaster@target was giving off a burnt smell.

(c) The cook@act in the kitchen@loc took the bread@obj out quickly, as the toaster@target was giving off a burnt smell.

(d) The man@act in the room@loc took the crumbs@obj out quickly, as the toaster@target was giving off a burnt smell.

(e) The bread@obj was taken out quickly by the the cook@act in the kitchen@loc, as the toaster@target was giving off a burnt smell.

(f) The crumbs@obj were taken out quickly by the the man@act in the room@loc, as the toaster@target was giving off a burnt smell.

23. (a) At the station@loc the conductor@act inspected the hitches@obj, while the train@target was stationary.

(b) At the stop@loc the man@act inspected the lights@obj, while the train@target was stationary.

(c) The conductor@act at the station@loc inspected the hitches@obj, while the train@target was stationary.

(d) The man@act at the stop@loc inspected the lights@obj, while the train@target was stationary.

(e) The hitches@obj were being inspected by the conductor@act at the station@loc, while the train@target was stationary.

(f) The lights@obj were being inspected by the man@act at the stop@loc, while the train@target was stationary.

24. (a) In the stadium@loc the referee@act was holding the yellow_card@obj, while the whistle@target was dangling from his pocket.

(b) In the park@loc the man@act was holding the watch@obj, while the whistle@target was dangling from his pocket.

(c) The referee@act in the stadium@loc was holding the yellow_card@obj, while the whistle@target was dangling from his pocket.

(d) The man@act in the park@loc was holding the watch@obj, while the whistle@target was dangling from his pocket.

(e) The yellow_card@obj was being held by the referee@act in the stadium@loc, while the whistle@target was dangling from his pocket.

(f) The watch@obj was being held by the man@act in the park@loc, while the whistle@target was dangling from his pocket.

A.3 Linguistic Stimuli used in Experiments 5 and 6

We provide the linguistic stimuli used in Experiment 5 and Experiment 6 (see Chapter 7). We report only the conditions: neutral_context:target_word (a), target_context:target_word (b), neutral_context:competitor_word (c), and competitor_context:competitor_word (d). The experiments also included the conditions: target_context:competitor_word, and competitor_context:target_word. The contextual properties generated with Strudel are surrounded by “*” while the critical word is surrounded by “#”.

1. (a) First, the man agreed hesitantly, but then he thought about the #cannon# and realised that it was scary.

   (b) When the soldier was on the *turret*, he *fired* at the *enemy* with his #cannon# again and again.

   (c) First, the man agreed hesitantly, but then he thought about the #bomb# and realised that it was scary.

   (d) When the soldier was in the *aircraft*, he *destroyed* the *target* with the #bomb# in a moment.

2. (a) Eventually, the man got ready quickly, and then he saw the #car# and said that he would like to drive it.

   (b) While the man was *driving* on a rural street, the *engine* light came on and he *parked* his #car# immediately.

   (c) Eventually, the man got ready quickly, and then he saw the #scooter# and said that he would like to ride it.

   (d) While the man was *riding* over a *ramp*, he increased the *speed* of his #scooter# suddenly.

3. (a) Initially, the man talked constantly, but then he looked at the #caterpillar# and watched in silence.

   (b) While the man looked at the *butterflies* in the zoo, he saw a *leaf* eaten by a particular *species* of #caterpillars# completely.

   (c) Initially, the man talked constantly, but then he looked at the #gorilla# and watched in silence.

   (d) While the man was at the *zoo*, he looked at the *cage* with a *group* of #gorillas# for a long time.

4. (a) At first, the man continued slowly, but then he looked at the #cooker# and realised that it was broken.

   (b) When the man was preparing a meal in the *kitchenette*, he *heated* it in a *pan* and used the #cooker# carefully.

   (c) At first, the man continued slowly, but then he looked at the #ladle# and realised that it was broken.

(d) When the man was preparing a *soup* in the kitchen, he *poured* it in a *bowl* and used the #ladle# carefully.

5. (a) Initially, the man disagreed strongly, but then he looked at the #corn# and realised that it was unusable.

   (b) After he had set up the *mill*, the *farmer* *sowed* the #corn# in the field.

   (c) Initially, the man disagreed strongly, but then he looked at the #broccoli# and realised that it was unusable.

   (d) While the man was preparing a meal in the kitchen, he *chopped* the *florets* and *boiled* the #broccoli# for a couple of minutes.

6. (a) Eventually, the man looked around carefully, and then he spotted the #drum# and decided that it may be worthwhile staying for the concert.

   (b) While the man was at the *band* parade, he listen to the *rhythm* of a *bass* and a #drum# from afar.

   (c) Eventually, the man looked around carefully, and then he spotted the #saxophone# and decided to play it.

   (d) While the man went to the *jazz* bar, he listen the *sound* of a *clarinet* and a #saxophone# all night long.

7. (a) At first, the man laughed loudly, but then he saw the #elephant# and understood that it was dangerous.

   (b) While the man was crossing the *jungle*, he saw a *poacher* *capturing* an #elephant# ferociously.

   (c) At first, the man laughed loudly, but then he saw the #alligator# and understood that it was dangerous.

   (d) While the man was crossing the *swamp*, he saw a *hippo* attacking a *gigantic* #alligator# ferociously.

8. (a) Initially, the man nodded silently, but then he pointed at the #guitar# and said that it disturbed him.

   (b) While the man was listening to a *song* in his room, the musician *played* a *riff* with his #guitar# for a couple of minutes.

(c) Initially, the man nodded silently, but then he pointed at the #clarinet# and said that it disturbed him.

(d) While the man was at the concert hall for a *sonata*, the players performed a *duet* comprising a *saxophone* and a #clarinet# for an hour.

9. (a) Eventually, the man smiled somewhat, and then he noticed the #harp# and thought it was beautiful.

(b) While the man was at the concert hall, the player *accompanied* the flute by *plucking* the *strings* of a #harp# for an hour.

(c) Eventually, the man smiled somewhat, and then he noticed the *violin* and thought that it was beautiful.

(d) While the man was at the concert hall for a *sonata*, the players performed a *duet* comprising a *piano* and a #violin# for an hour.

10. (a) In the beginning, the man thought carefully, but then he spotted the #kettle# and realised that it was unusable.

(b) While the man was preparing some *tea*, he turned on the *stove* and *filled* the #kettle# repeatedly.

(c) In the beginning, the man thought carefully, but then he spotted the #dishwasher# and realised that it was unusable.

(d) While the man was tidying up the *kitchen*, he took the *detergent* and *used* the #dishwasher# repeatedly.

11. (a) At first, the woman looked confused, but then she saw the #mitten# and agreed that it was ugly.

(b) The man was walking and his *fingers* froze, so he *warmed* up and *wore* his #mittens# immediately.

(c) At first, the woman looked confused, but then she saw the #scarf# and agreed that it was ugly.

(d) The man was walking and his *neck* was hurting, so he *covered* it and *put* on his #scarf# immediately.

12. (a) First, the man worried greatly, but then he saw the #peacock# and realised that it was fine.

(b) While the man was crossing the *garden*, he saw the *feathered* *tail* of a #peacock# for a couple of seconds.

(c) First, the man worried greatly, but then he saw the #owl# and recognised that it was fine.

(d) While the man was crossing the forest in the *night*, he *heard* the *hoot* of an #owl# for a couple of minutes.

13. (a) At first, the man turned away for a short time, but then he thought about the #piano# and realised that it must be expensive.

(b) While the man was at the playhouse for a *sonata*, a musician *accompanied* the *guitar* with his #piano# for the rest of the evening.

(c) At first, the man turned away for a short time, but then he thought about the #cello# and realised that it must be expensive.

(d) While the man was at the playhouse for a *concerto*, the players performed a *duet* comprising a *viola* and a #cello# for an hour.

14. (a) At first, the woman agreed cheerfully, but then she saw the #potato# and spotted that it was mouldy.

(b) While the man was preparing a meal in the kitchen, he took the *onions* and he *peeled* and *mashed* the #potatoes# slowly.

(c) At first, the woman agreed cheerfully, but then she saw the #cucumber# and spotted that it was mouldy.

(d) While the man was preparing a *sandwich* in the kitchen, he took the *yoghurt* and *sliced* the #cucumber# quickly.

15. (a) First, the man nodded quickly, but then he looked at the #refrigerator# and realised that it was broken.

(b) When the man went to the *kitchen*, he *kept* some *beers* and plugged in the #refrigerator# for later.

(c) First, the man nodded quickly, but then he looked at the #mixer# and realised that it was broken.

(d) When the man was preparing some *drinks*, he took a *bowl* and he *installed* the #mixer# for later.

16. (a) First, the man disagreed somewhat, but then he noticed the #toaster# and appreciated that it was useful.
    (b) When the man was preparing his breakfast, he *got* some *bread* and *used* the #toaster# immediately.
    (c) First, the man disagreed somewhat, but then he noticed the #corkscrew# and appreciated that it was useful.
    (d) When the man was organising the *wine* tasting, he *bought* some *bottles* and a new #corkscrew# for later.

17. (a) After that the man turned away gradually, and then he noticed the #train# and said that it was a noisy place.
    (b) While the man was at the *station* consulting the *timetable*, the *passengers* heard the horn of the #train# from afar.
    (c) After that the man turned away gradually, and then he noticed the #plane# and said that it was a noisy place.
    (d) While the man was waiting to *board*, he saw the *pilot* *manoeuvring* his #aeroplane# with difficulty.

18. (a) Initially, the man disagreed somewhat, but then he looked at the #waistcoat# and realised that it was ugly.
    (b) While the cowboy was going to the bar, he *buttoned* up the *shirt* and *wore* his favourite #waistcoat# for the rest of the evening.
    (c) Initially, the man disagreed somewhat, but then he looked at the #trousers# and realised that they were ugly.
    (d) The man was going out, so he *wore* his *socks*, *shirt*, and #trousers# for the evening.

References

Almuhareb, A., & Poesio, M. (2005). Concept learning and categorization from the web. In *Proceedings of the 27th annual conference of the cognitive science society* (pp. 103–108). Stresa, Italy.

Altmann, G. T. M. (1999). Thematic role assignment in context. *Journal of Memory and Language*, *41*(1), 124–145.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: the 'blank screen paradigm'. *Cognition*, *93*(2), B79–B87.

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

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238.

Anderson, R. C., Pichert, J. W., Goetz, E. T., Schallert, D. L., Stevens, K. V., & Trollip, S. R. (1976). Instantiation of general terms. *Journal of Verbal Learning and Verbal Behavior*, *15*(6), 667–679.

Andrews, M., Vigliocco, G., & Vinson, D. P. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological review*, *116*(3), 463–498.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora. *Language Resources and Evaluation*, *43*(3), 209–231.

Baroni, M., & Lenci, A. (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, *36*(4), 673–721.

Baroni, M., Murphy, B., Barbu, E., & Poesio, M. (2010). Strudel: A Corpus-Based Semantic Model Based on Properties and Types. *Cognitive Science*, *34*(2), 222–254.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 1–51.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of memory and language*, *63*(4), 489–505.

Blacoe, W., & Lapata, M. (2012). A Comparison of Vector-based Representations for Semantic Composition. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning.* (pp. 546–556).

Clark, R. A., Richmond, K., & King, S. (2007). Multisyn: Open-domain unit selection for the Festival speech synthesis system. *Speech Communication*, *49*, 317–330.

Coco, M. I., & Keller, F. (2012). Scan Patterns predict Sentence Production in the Cross-modal Processing of Visual Scenes. *Cognitive Science*, *in press*.

Conrad, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory & Cognition*, *2*(1A), 130–138.

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*(2), 163–201.

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.

Devereux, B. J., Pilkington, N., Poibeau, T., & Korhonen, A. (2010). Large-Scale Acquisition of Feature-Based Conceptual Representations from Textual Corpora. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *7*, 137–170.

Duffy, S. A., Henderson, J. M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 791–801.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior*, *20*, 641–655.

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505.

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495.

Federmeier, K. D., McLennan, D. B., De Ochoa-Dewald, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study. *Psychophysiology*, *39*(2), 133–146.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*(75), 75–84.

Ferreira, F., & Henderson, J. (1990). Use of Verb Information in Syntactic Parsing: Evidence From Eye Movements and Word-by-Word Self-Paced Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 555–568.

Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., et al. (2002). Placing Search in Context: The Concept Revisited. *ACM Transactions on Information Systems*, *20*(1), 116–131.

Firth, J. R. (1957). A synopsis of language theory 1930-1955. In *Studies in linguistic analysis* (pp. 1–32). Oxford, UK: Blackwell Publishers.

Fischler, I. S., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, *18*, 1–20.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Foote, R. (2010). Integrated knowledge of agreement in early and late EnglishSpanish bilinguals. *Applied Psycholinguistics*, *32*(01), 187–220.

Frassinelli, D., & Keller, F. (2012). The Plausibility of Semantic Properties Generated by a Distributional Model: Evidence from a Visual World Experiment. In *Proceedings of the 34th annual conference of the cognitive science society* (pp. 1560–1565). Sapporo.

Frassinelli, D., Keller, F., & Scheepers, C. (2013). The Effect of Incremental Context on Conceptual Processing : Evidence from Visual World and Reading Experiments.

In *Proceedings of the 35th annual conference of the cognitive science society* (pp. 460–466). Berlin.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, *11*(4), 274–279.

Hare, M., Jordan, M. I., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, *111*(2), 151–167.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(23), 146–162.

Hofmeister, P. (2011). Representational Complexity and Memory Retrieval in Language Comprehension. *Language and cognitive processes*, *26*(3), 376–405.

Hogaboam, T. W., & Perfetti, C. A. (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *14*, 265–274.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University 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*(1), B23–32.

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*, *57*(4), 460–482.

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. M. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta psychologica*, *121*(1), 65–80.

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

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813.

Hwang, A. D., Wang, H.-C., & Pomplun, M. (2009). Semantic guidance of eye movements during real-world scene inspection. *Journal of Vision*, *9*(8), 2534–2539.

Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.

Jegerski, J., & VanPatten, B. (2013). *Psycholinguistics and Second Language Acquisition*. Routledge.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(04), 603–634.

Johnson-Laird, P. N. (1987). The mental representation of the meaning of words.

Cognition, *25*, 189–211.

Jones, M. N., Kintsch, W., & Mewhort, D. J. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552.

Jordan, T. R., & Thomas, S. M. (2002). In search of perceptual influences of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 34–45.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238.

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*(1), 133–156.

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. *The computer and literary studies*, 153–165.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in cognitive sciences*, *4*(12), 463–470.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(12), 161–163.

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. *Preparatory states and processes*, 217–237.

Landauer, T., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211–240.

Lapesa, G., & Evert, S. (2013). Evaluating Neighbor Rank and Distance Measures as Predictors of Semantic Priming. In *Proceedings of the workshop on cognitive modeling and computational linguistics* (pp. 66–74).

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Rivista di Linguistica*, *20*(1), 1–31.

Lin, D. (1997). Using syntactic dependency as local context to resolve word sense ambiguity. *Proceedings of the 35th annual meeting on Association for Computational Linguistics*, 64–71.

Lucas, M. M. (1987). Frequency effects on the processing of ambiguous words in sentence contexts. *Language and Speech*, *30*, 25–46.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods Instruments and Computers*, *28*, 203–208.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology*, *10*, 29–63.

McClelland, J. L., & Elman, J. L. (1986). The TRACE Model of Speech Perception. *Cognitive Psychology*, *18*, 1–86.

McClelland, J. L., & O'Regan, J. (1981). Expectations increase the benefit derived from parafoveal visual information in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 634–644.

McClelland, J. L., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review*, *88*(5), 375–407.

McDonald, S. A. (2000). *Environmental determinants of lexical processing effort*. Unpublished doctoral dissertation, University of Edinburgh.

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14*(6), 648–652.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 558–572.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods*, *37*(4), 547–559.

McRae, K., Hare, M., Elman, J. L., & Todd, F. (2005). A basis for generating expectancies for verbs from nouns. , *33*(7), 1174–1184.

McRae, K., Sa, V. R. de, & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*(2), 99–130.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the Influence of Thematic Fit (and Other Constraints) in On-line Sentence Comprehension. *Journal of Memory and Language*, *38*(3), 283–312.

Medin, D. L. (1989). Concepts and conceptual structure. *The American psychologist*, *44*(12), 1469–81.

Miller, G. A. (1978). Practical and lexical knowledge. In *Cognition and categorization* (pp. 305–320). Hillsdale, NJ: Erlbaum.

Mitchell, J. (2011). *Composition in Distributional Models of Semantics*. Unpublished doctoral dissertation, University of Edinburgh.

Mitchell, J., & Lapata, M. (2008). Vector-based models of semantic composition. *proceedings of ACL-08: HLT*(June), 236–244.

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 92–103.

Moss, H. E. (1997). The Time Course of Activation of Semantic Information during Spoken Word Recognition. *Language and Cognitive Processes*, *12*(5-6), 695–732.

Moss, H. E., & Marslen-Wilson, W. D. (1993). Access to word meanings during spoken language comprehension: effects of sentential semantic context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1254–76.

Moss, H. E., & Older, L. (1996). *Birkbeck word association norms*. Psychology Press.

Murphy, G. L. (2002). *The Big Book of Concepts*. London: MIT Press.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in cognitive sciences*, *11*(12), 520–27.

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, *9*(3), 225–236.

Padó, S., & Lapata, M. (2007). Dependency-based construction of semantic space models. *Computational Linguistics*, *33*(2), 161–199.

Peirsman, Y., Heylen, K., & Speelman, D. (2008). Putting things in order. First and second order context models for the calculation of semantic similarity. *9es Journées internationales d'Analyse statistique des Données Textuelles*, 907–916.

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number-of-features effects in lexical decision and naming tasks. *Psychonomic bulletin & review*, *9*(3), 542–549.

Posner, M. I., Rafal, R. D., Choate, L. S., & Jonathan, V. (1985). Inhibition of return : Neural basis and function Inhibition of Return : Neural Basis and Function. *Cognitive Neuropsychology*, *2*(3), 211–228.

Prather, P., & Swinney, D. A. (1988). Lexical processing and ambiguity resolution: An autonomous process in an interactive box. In *Lexical ambiguity resolution* (pp.

289–310). New York: Morgan Kaufman.

Rayner, K., & Clifton, C. (2002). Language Processing. In *Handbook of experimental psychology* (pp. 261–316). New York: John Wiley and Sons.

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 779–790.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic bulletin & review*, *3*(4), 504–509.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, *7*(4), 573–605.

Sahlgren, M. (2006). *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Unpublished doctoral dissertation, Stockholm University.

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics*, *20*(1), 33–53.

Saussure, F. de. (1915). *Cour de Linguistique Générale (Course in General Linguistics)*. Trans. Wade Baskin (1959 edition).

Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics*, *24*(1), 97–123.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 344–354.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The Influence of Sentence Constraint for Upcoming on the Scope of Facilitation Words. *Journal of Memory and Language*, *24*, 232–252.

Sedivy, J. C., Tanenhaus, M. K., & Chambers, C. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, *14*(4), 489–537.

Sereno, S. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal* *of experimental psychology. Human learning and memory*, *6*(2), 174–215.

Stanovich, K., & West, R. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, *112*, 1–36.

Staub, A., Abbott, M., & Bogartz, R. S. (2012). Linguistically guided anticipatory eye movements in scene viewing. *Visual Cognition*, *20*(8), 922–946.

Swaab, T., Brown, C. M., & Hagoort, P. (2003). Understanding words in sentence contexts: The time course of ambiguity resolution. *Brain and Language*, *86*(2), 326–343.

Swinney, D. A., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory & Cognition*, *7*(3), 159–175.

Tabossi, P. (1988). Effects of context on the immediate interpretation of unambiguous nouns. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 153–162.

Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity : Effects of context and dominance. *Psychological Research*, *49*, 161–167.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 427–440.

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*, 1632–34.

Taylor, W. L. (1953). " Cloze procedure": a new tool for measuring readability. *Journalism Quarterly*, *30*, 415–433.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review*, *113*(4), 766–86.

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1266–1282.

Trueswell, J. C., & Kim, A. E. (1998). How to Prune a Garden Path by Nipping It in the Bud: Fast Priming of Verb Argument Structure. *Journal of Memory and Language*, *39*(1), 102–123.

Turney, P. D., & Pantel, P. (2010). From Frequency to Meaning : Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467.

Van Petten, C., & Kutas, M. (1988). Tracking the time course of meaning activation. In *Lexical ambiguity resolution* (pp. 431–475).

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, *83*(2), 176–90.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*(4), 422–488.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92–114.

Wicha, N., Moreno, E., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288.

Williams, J. (1988). Constraints upon semantic activation during sentence comprehension. *Language and Cognitive Processes*, *3*, 165–206.

Wittgenstein, L. (1953). *Philosophical investigations.* Oxford, UK: Blackwell Publishers.

Wu, L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: evidence from property generation. *Acta psychologica*, *132*(2), 173–89.

Yee, E., Overton, E., & Thompson-Schill, S. L. (2009). Looking for meaning: eye movements are sensitive to overlapping semantic features, not association. *Psychonomic bulletin & review*, *16*(5), 869–74.