The comprehension benefit of noun-contingent eye movements in the visual world

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultäten
der Universität des Saarlandes

vorgelegt von

Juliane Steinberg

aus Halle/Saale

Saarbrücken, 2012

Dekan:	Prof. Dr. Roland Marti
Berichterstatter:	Prof. Dr. Matthew W. Crocker
	Prof. Dr. Bernd Möbius
Tag der letzten Prüfungsleistung:	02.05.2012

Abstract

When we listen to someone talking, we try, at any given point in time, to make the connection between what is being said and the real world. Language processing incorporates information from various different sources and modalities in order to create an immediate and adequate representation of the conveyed meaning. Language processing also affects the information processing in other modalities, specifically visual perception: looks to entities in one's visual environment while listening to spoken utterances closely mirror one's comprehension of referring expressions in the speech stream. This behavior - noun-contingent eye movements - has been observed consistently in psycholinguistic research and has been used as a window into the conceptual layer of the language comprehension process, shedding light onto an extremely wide range of language processing phenomena. The phenomenon of noun-contingent eye movements itself is not yet fully understood though. Drawing back on concepts from spatial indexing, lexical access and priming we motivate the search for and investigation of a beneficial effect of noun-contingent referent looks on the noun's speed of comprehension.

This thesis utilizes a novel experimental methodology which makes it possible to simultaneously observe and manipulate participants' inclination to make noun-contingent eye movements as well as record the speed of comprehension for the spoken noun that would trigger these eye movements. This is necessary if one wants to investigate a potential effect of eye movements on noun comprehension. The methodology is a combination of a classical visual-world setting, where participants listen to single spoken nouns and inspect an array of objects that may be related to the noun, and of a lexical decision task which is performed on that spoken noun and assesses the speed of its recognition.

With the experiments presented in this thesis we manage to provide reliable empirical evidence for a beneficial effect of noun-contingent referent looks on the speed of noun comprehension. It is also shown that this benefit arises from the perception of the referent's visual image (as opposed to the activation of internal referent representations or avoidance of interference from unrelated visual images) during the comprehension of the spoken noun (as opposed to a benefit from the initial recognition and encoding of the referent object). Finally, it is proposed that a spatial-indexing mechanism plays a role in mediating the strength of the effect that the

visual image provided by a noun-contingent eye movement can have on the comprehension of the spoken noun.

Zusammenfassung

Die Verarbeitung von sprachlichen und bildlichen Reizen ist beim Menschen eng miteinander verwoben. Wer gleichzeitig gesprochener Sprache zuhört und dazu passende Dinge in seiner Umgebung sieht, wird häufig seinen Blick davon steuern lassen, welche Worte er gerade versteht. Das heißt beispielsweise, dass man, während man eine Objektbezeichnung hört, seinen Blick auf das benannte Objekt richtet. In einer Situation, in der auf ein Objekt gezeigt wird ("Schau mal, der Apfel!"), ist die Motivation für solch ein Verhalten intuitiv verständlich. Aber auch in Situationen, in denen sich gesprochene Äußerungen nicht unmittelbar auf die Umgebung beziehen, blicken Zuhörer verlässlich auf Objekte, die zu den gerade geäusserten Worten passen. Solche Blicke sind nicht unmittelbar pragmatisch motiviert, möglicherweise aber dadurch, dass das *Sprachverstehen* von diesem Blickverhalten profitiert. Dieser Möglichkeit gehen wir in der vorliegenden Arbeit auf den Grund.

Hierzu nutzen wir eine neuartige Kombination von Forschungsmethoden, mit der man gleichzeitig Einblicke sowohl in das Sprachverstehen als auch in die visuelle Verarbeitung gewinnt. Mit dem 'Visual World'-Paradigma werden in der psycholinguistischen Forschung Erkenntnisse über zahlreiche Aspekte der Sprachverarbeitung gewonnen: In Studien zum Sprachverstehen wird der Blick eines Zuhörers zu einem Objekt, das in einer Beziehung zu einem zeitnah geäußerten Wort steht, als Zeichen dafür verstanden, dass der Zuhörer dem Wort (wenigstens zeitweise) die Bedeutung dieses Objekts zuordnet. Indem man nun mittels Eyetracking beobachtet, welche Objekte in Ihrer Umgebung Menschen anschauen, während sie gesprochene Sprache verstehen, erfährt man in einer hohen zeitlichen Auflösung, wie gesprochene Äußerungen interpretiert werden. Mit Studien dieser Art werden Blickbewegungen als Konsequenz der Sprachverarbeitung genutzt; man kann also erfahren, wie das Verstehen gesprochener Sprache die Blickbewegungen steuert. Die zweite Seite unserer zentralen Forschungsfrage ist die nach dem Einfluss der visuellen auf die sprachliche Verarbeitung. Wie wird das Verstehen eines Wortes davon beeinflußt, was man gerade sieht? Solchen Fragestellungen widmen sich Priming- und Lexical-Decision-Studien. Im Priming-Paradigma wird mit verschiedenartigen Aufgaben erforscht, wie sich die Verarbeitung eines Konzept (des Targets) durch den Kontext eines weiteren Konzepts (des Primes) verändert: Indem man Probanden beispielsweise zusätzlich zu einem

zu benennenden Objekt ein weiteres Bild zeigt, das in einer engen Beziehung zu dem zu benennenden Objekt steht, gewinnt man Aufschluß darüber, wie ähnliche bildliche Informationen den Benennensprozess beeinflussen. Die lexikalische Entscheidungsaufgabe (Lexical Decision) besteht darin, zu einer kurzen Laut-oder Buchstabenfolge zu entscheiden, ob es sich dabei um ein echtes Wort handelt. Auf unsere Fragestellung angewandt ermöglichen es diese Methoden, herauszufinden, wie Blicke zu verschiedenen Arten von Objekten (Primes) das Verstehen eines gesprochenen Wortes (Targets), erfasst durch die lexikalische Entscheidungsaufgabe, beeinflussen.

Aus der Blickbewegungsforschung mit der 'Visual World'-Methode und aus zahlreichen Primingstudien stammt auch das Wissen über die Phänomene, auf denen unsere Forschungsfragen fußen: Wie bereits erwähnt, schaut man häufig Objekte in seiner Umgebung an, während man Äußerungen versteht, die sich direkt oder indirekt auf diese Objekte beziehen. Diese Beobachtung wurde zuerst von Roger Cooper gemacht (Cooper, 1974) und seit der Studie von Tanenhaus, Spivey-Knowlton, Eberhard und Sedivy (1995) in zahlreichen Untersuchungen zu verschiedensten Details des Sprachverstehensprozesses repliziert. Gesprochene Sprache wird zu jedem Zeitpunkt so weit verarbeitet, wie es die bis dahin wahrgenommenen Laute es erlauben. Das heißt, dass neben dem Referenten eines gesprochenen Wortes (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Magnuson, Dixon, Tanenhaus, & Aslin, 2007) anfänglich auch Objekte angeschaut werden, deren Name mit den gleichen Phonemen beginnt, selbst wenn ihre Bedeutung sich von dem gesprochenen Wort unterscheidet (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan & Gaskell, 2007). Auch nicht ähnlich klingende Objekte werden fixiert, wenn ihr Sinn dem des Wortes ähnelt (Yee & Sedivy, 2001; Huettig & Altmann, 2005). So bildet das Blickverhalten die Aktivierung all derjenigen internen Repräsentationen ab, die bei der Zuordnung von Bedeutung zu einer wahrgenommenen Lautfolge in Betracht gezogen werden. Die drei Verwandtschaftsbeziehungen Referenz, phonologische Ähnlichkeit und semantische Ähnlichkeit spielen auch in Primingstudien eine Rolle. Wenn der Prime-Stimulus vor dem Target präsentiert wird, können grundsätzlich entgegengesetzte Effekte beobachtet werden: Referenten und semantisch ähnliche Primes beschleunigen die Verarbeitung eines Target-Konzepts (Levelt, Vorberg, Meyer, Pechmann, & Havinga, 1991; McQueen & Huettig, 2005). Phonologisch ähnliche Objekte dagegen haben einen eher nachteiligen Effekt auf die Verarbeitung eines danach präsentierten Target-Konzepts (Goldinger, Luce, Pisoni, & Marcario, 1992; Levelt et al., 1991). Somit gibt es drei verschiedene Arten von Objekten, die während des Verstehens einer Objektbezeichnung verlässlich fixiert werden, die aber unterschiedliche Auswirkungen auf den Verstehensprozess haben sollten.

Unsere ersten beiden Experimente dienen als Einstieg in die Frage, wie das Verstehen von

Objektbezeichnungen und die Steuerung von Blicken zu möglichen Referenten miteinander interagieren. Um eine Situation herzustellen, die vergleichbar mit anderen Studien ist, in denen Blickbewegungen durch das Verstehen von gesprochenen Objektbezeichnungen ausgelöst werden, sahen die Probanden vier (bzw., im zweiten Experiment, drei) Objekte auf einem Computerbildschirm und hörten nach einer kurzen Zeitspanne dazu ein gesprochenes Wort. Die Identität von einem der gezeigten Objekte wurde experimentell variiert: Dieses Objekt konnte entweder dem gesprochenen Wort entsprechen (benanntes Objekt, Referent), sein Name konnte mit den gleichen Phonemen beginnen wie das gesprochene Wort (phonologisch ähnliches Objekt), es konnte der Bedeutung des gesprochenen Wortes ähneln (semantisch ähnliches Objekt), oder es konnte ein Objekt völlig ohne Bezug zu dem gesprochenen Wort sein (nicht verwandtes Objekt). Die Probanden hatten die Aufgabe, zu entscheiden, ob die gesprochenen Laute ein Wort waren oder nicht. So sollten Experiment 1 und 2 erste Hinweise darauf geben, ob die Tendenz, mit einem gesprochenen Wort verwandte Bilder anzuschauen, davon abhängt, welchen Einfluss diese Bilder auf das Sprachverstehen haben können (siehe oben: Referent und semantisch ähnliches Objekt sollten hilfreich sein, während das phonologisch ähnliche Objekt das Sprachverstehen behindern sollte). Ebenso sollte gezeigt werden, ob der Einfluss der dargestellten Bilder auf das Sprachverstehen, der bisher nur aus Primingstudien bekannt war, überhaupt in einer Visual-World-Situation bestehen würde. Die Ergebnisse der beiden Experimente zeigen zunächst, dass alle drei Arten von dem gesprochenen Wort verwandten Objekten während des Sprachverstehens angeschaut werden. Zudem beschleunigt es klar die lexikale Entscheidung, den Referenten des gesprochenen Wortes zu sehen. Wir finden keinen Effekt von semantisch ähnlichen Objekten auf die lexikale Entscheidung, und die Ergebnisse in bezug auf das phonologisch ähnliche Objekt bleiben zunächst inkohärent.

Experiment 3 fügt der bisherigen offenen Fragestellung den Aspekt der Kausalität hinzu: Welche Auswirkungen haben Blicke auf verwandte Objekte während des Verstehens eines gesprochenen Wortes auf den Verstehensprozess? Um diese Frage beantworten zu können, dürfen die Blickbewegungen nicht, wie normalerweise, durch das Sprachverstehen gesteuert werden, sondern müssen experimentell manipuliert werden. Zusätzlich zu dem Faktor des Bezugs von einem der dargestellten Objekte zu dem gesprochenen Wort wird in Experiment 3 deshalb variiert, welches Objekt mit dem Beginn des gesprochenen Wortes visuell hervorgehoben wird. Durch diesen zusätzlichen Reiz werden die Blicke während des gesprochenen Wortes entweder auf das manchmal verwandte Objekt (s.o.) oder auf ein anderes, immer nicht verwandtes Objekt gezogen. Wir können so zeigen, dass Blicke auf den Referenten *während des gesprochenen Wortes* das Verstehen des gesprochenen Wortes erleichtern. Interessanterweise genügt es nicht, den Referenten *vor dem Hören des Wortes* gesehen zu haben, wenn man ihn während der Lebensdauer des Wortes nicht anschaut. Semantisch ähnliche Objekte zeigen weiterhin keinen Effekt, Blicke auf phonologisch ähnliche Objekte während des gesprochenen Wortes verursachen dagegen gehäuft Fehler in der lexikalen Entscheidung.

Als Kontrollexperiment bestätigt Experiment 4 unter anderem, dass mit unseren Stimuli eine Interferenz von phonologisch ähnlichen Objekten mit dem Verstehen des gesprochenen Wortes möglich ist, dass wir aber keinen Effekt von unseren semantisch ähnlichen Objekte auf die lexikale Entscheidung sehen können. In diesem Experiment wurden die betreffenden Objekte einzeln auf dem Bildschirm kurz vor und während der lexikalen Entscheidungsaufgabe präsentiert.

Mit den ersten vier Experimenten haben wir somit gezeigt, dass die Blickbewegungen während des Sprachverstehens eine Auswirkung auf das Sprachverstehen haben. Besonders deutlich ist es, dass man davon profitiert, benannte Objekte anzuschauen, während man deren Namen hört. Die nachfolgenden Experimente beschäftigen sich tiefergehend mit der Frage, wie dieser Vorteil durch Blicke auf den Referenten des gesprochenen Wortes entsteht.

Im anschließenden Experiment 5 wird näher untersucht, ob verschiedene Arten von Priming-ähnlichen Effekten den Vorteil von Blicken auf Referenten für das Sprachverstehen erklären können. Indem manipuliert wird, ob Versuchspersonen unter den gezeigten Bildern einen Referenten für das gesprochene Wort finden können, und wohin mit Beginn des gesprochenen Wortes mittels eines visuellen Reizes der Blick gelenkt wird (auf den Referenten, auf eine leere Stelle des Bildschirms oder auf ein nicht verwandtes Objekt), wird zwischen drei Hypothesen entschieden. Eine Möglichkeit ist die, dass die Blicke auf den Referenten während des gesprochenen Wortes lediglich dem Erhalten eines Priming-Effektes dienen, der schon daraus erwächst, dass man den Referenten gesehen und registriert hat, bevor man das gesprochene Wort hört. Wenn Probanden jedoch den Referenten zwar vor, aber nicht mehr während des gesprochenen Wortes anschauen konnten, verstanden sie das Wort genauso langsam wie wenn sie den Referenten nie gesehen hatten. Somit finden wir für diese Hypothese keine Bestätigung. Außerdem erschwert es das Verstehen des Namens nicht, auf ein nicht verwandtes Objekt zu schauen (im Vergleich dazu, kein Objekt anzuschauen), daher ist es auch unwahrscheinlich, dass das Sprachverstehen davon profitiert, nicht verwandte Objekte zu meiden. Klar zeigt dieses Experiment dagegen wieder, dass es hilfreich ist, *während* des Sprachverstehens auf den Referenten des gesprochenen Wortes zu schauen, man profitiert also unmittelbar von dieser Blickbewegung oder der Information, die sie zur Verfügung stellt.

Die letzten beiden Experimente dieser Arbeit ergründen die Rolle von an den Ort der gesehenen Objekte gebundener Information. Im Rahmen von Theorien über Spatial Indexing und External Memory (Spivey, Richardson, & Fitneva, 2004) wird angenommen, dass man auf Objekte in seiner visuellen Umgebung jederzeit aufgabenbasiert und effektiv zugreift, indem man sie über Indizes in der Umwelt verortet und dem jeweiligen Aufgabenkontext entsprechend intern benennt. Auch Blickbewegungen während des Sprachverstehens können so erklärt werden. Wenn das beginnende Verstehen eines gesprochenen Wortes zu der Benennung vorhandener Indizes passt, werden diese aktiviert, mit ihnen die assoziierten Positionen im Raum, und dadurch werden Blicke zu diesen Positionen ausgelöst. Über dieses Verständnis hinaus kann man die Hypothese aufstellen, dass ein Blick zu solch einer indizierten Position im Raum auch Zugang zu assoziierten internen Repräsentationen des dort gesehen Objekts gibt, deren Aktivierung dem Verstehen des gesprochenen Wortes zuträglich sein könnte. Experiment 6 prüft diese Hypothese, indem den Probanden zunächst ein Bildschirm mit drei Objekten gezeigt wird, von denen eines der Referent des gesprochenen Wortes ist. Gleichzeitig mit dem Beginn des Wortes werden jedoch alle drei Objekte ausgeblendet, lediglich ein leerer Kreis lenkt den Blick auf entweder die Position, in der vorher der Referent zu sehen war oder auf die vorherige Position eines nicht verwandten Objektes. Die Geschwindigkeit, mit der das gesprochene Wort verstanden wurde, unterschied sich zwischen diesen beiden Bedingungen nicht; wir sehen also keinen Beleg dafür, dass Blicke auf die bloße Stelle, an der vorher ein Referent gesehen wurde, der Sprachverarbeitung hilfreiche Informationen zur Verfügung stellen.

Womöglich spielen die Indizierungsmechanismen dennoch eine Rolle dabei, dass Blicke zu Referenten das Verstehen einer gesprochenen Objektbezeichnung beschleunigen. Experiment 7 stellt die Beziehung des gesehenen Objekts und dessen Position während des Wortverstehens einander gegenüber, um dies zu klären. Ein Referent und zwei nicht verwandte Objekte wurden gezeigt, und mit dem Beginn des gesprochenen Wortes verblieb nur noch ein einzelnes Objekt auf dem Bildschirm. Dies war entweder der Referent oder ein nicht verwandtes Objekt, und es konnte entweder in der vorherigen Position des Referenten oder der des nicht verwandten Objektes gezeigt werden. Wenn während der Lebensdauer des gesprochenen Wortes der Referent gezeigt wurde, beschleunigte er das Verstehen, egal wo er sich befand. Wurde das nicht verwandte Objekt gezeigt, war die lexikale Entscheidung deutlich langsamer wenn es sich in der vorherigen Position des Referenten befand als in seiner eigenen vorherigen Position. Aus diesen Ergebnissen können wir zunächst folgern, dass offensichtlich der Blick zu der Position, an der vorher ein Referent gesehen wurde, Information aktiviert, die mit dem Sprachverstehen interagiert. Dies ist nur der Fall, wenn an dieser fixierten Position ein Objekt gesehen wird, wobei dieses Objekt nicht der Referent sein muss. Wir interpretieren die Rolle dieser Information wie folgt: die Aktivierung eines Index, der zu einem gesprochenen Wort passt, ermöglicht es erst, dass die indizierte visuelle Information, die über eine Fixation aufgenommen wird, mit dem Verstehensprozess interagiert. Verweist der aktivierte Index wie erwartet auf passende visuelle

Information, profitiert das Sprachverstehen, befindet sich an dieser Stelle jedoch ein unpassendes Objekt, kann es zu Interferenz kommen.

Wir finden mit dieser Arbeit in mehreren Experimenten Belege dafür, dass Blickbewegungen zu verwandten Objekten beim Verstehen einer Objektbezeichnung eine Auswirkung auf die Geschwindigkeit des Sprachverstehens haben. Schaut man während des Verstehens eines Wortes auf seinen Referenten, kann das den Verstehensprozess beschleunigen. Dabei profitiert man direkt davon, zeitgleich mit dem Verstehensprozess passende visuelle Informationen aufzunehmen. Den 'Spatial Indices' in diesem Verhältnis kommt vermutlich eine regulierende Rolle zu: nur, wenn die Fixation eines Objekts durch die Aktivierung eines zum gesprochenen Wort passenden Indizes ausgelöst wird, kann die aufgenommene visuelle Information mit dem Verstehen des Wortes interagieren.

Acknowledgements

It was a long journey here, and there are some people that I need to thank for being there along the way:

First, I would like to thank Matt Crocker for supervising this work, discussing methods and designs for hours and hours and for giving me the opportunity of being a scientist for a while.

I've met lots of colleages and friends through this work who gave me countless bits of valuable advice and helped by being friends: Thank you, Alissa, Andrea, Sabine, Kerstin, Jens, Judith, Emilia, Maria, Ulrike, Garance, Afra, Pirita.

Mark, thank you for being a friend in everyday office life.

Und zuletzt:

Danke, Berry, für das unermüdliche und entscheidende Aufbauen.
Danke, Mami, für den Blick aus der Realität.
Danke, Basti, dafür, dass Du bis zum Ende mit mir durchgehalten hast.

Contents

List of Figures

List of Tables

1 Background

1.1 Introduction

Language use is often visually situated: While people produce or comprehend spoken language, what they see is often related to what is being talked about. Traditionally, psycholinguistics has idealized away from the situatedness and focussed on comprehension in 'isolation'. Over the past 15 years, however, there has been an increase in experimental evidence emphasizing that lexical processes are grounded in the visual environment through visual attention (Tanenhaus et al., 1995; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Kamide, Altmann, & Haywood, 2003; Knoeferle & Crocker, 2007). This thesis seeks to understand why this is the case, by examining the influence of the direction of visual attention on spoken word comprehension.

In the absence of a visible conversation partner, people have a consistent tendency to look at objects they talk about or hear someone talk about, in close temporal proximity to those objects being mentioned. While talking about something he sees, a speaker's gaze in his environment often reflects his own speech planning: While planning the structure of his utterance, he will look at the entities that he considers in this process and he will fixate entities shortly before naming them out loud (Griffin & Bock, 2000; Griffin, 2004). A listener, on the other hand, will likely direct his eyes towards visually present objects that are mentioned by the speaker. He will often begin to look at mentioned entities almost immediately after the speech stream identifies them - more specifically, at entities that are potential referents of the speech stream given the information received so far (Tanenhaus et al., 1995; Magnuson et al., 2007; Altmann & Kamide, 1999). In addition, listeners often look at some entities related to the referred-to objects, at entities that might plausibly be referred to soon and at entities relevant to the described situation and actions.

The focus of this thesis is on the listeners and their looks to objects in response to object nouns in the speech stream, that is on listeners' so-called noun-contingent eye movements. These noun-contingent eye movements (NCEs) are made consistently, even when people know the entities in the scene and when the scene is not explicitly relevant to the utterance. Listeners take the effort to match the focus of their visual attention in the shared visual environment closely to what they believe the speaker is referring to. Why would they do this? Why does the sensomotoric behavior in the visual modality closely track or mirror the processes of language comprehension, even in cases where the perceived visual information does not provide significant additional information for the interpretation of the comprehended utterance? The present thesis investigates the idea that this eye movement behavior might not only be caused by, but also contribute to language processing, specifically by facilitating the word recognition process. In the studies described below, we explore the existence of such a potential benefit and examine the mechanisms which may underlie it.

There are two groups of established behavioral phenomena in psycholinguistic research which our studies build upon: noun-contingent eye movements to objects related to a comprehended noun, as found in studies employing the visual-world paradigm (Tanenhaus et al., 1995; Allopenna et al., 1998; Yee & Sedivy, 2001), and the effects that the viewing of an object can have on the comprehension of a concurrently perceived related noun, as investigated in priming and auditory lexical decision studies (Levelt et al., 1991; Goldinger, 1996). We develop a novel experimental setting that combines these two methods in order to explore the bi-directional interaction between visual and linguistic processing.

In the following sections we will first briefly introduce the research paradigms we combined in our own experimental paradigm. This is followed by a description of the behavioral patterns our hypotheses are based on: how likely are people to look at objects related to spoken nouns they hear, and how much does the processing of one concept prime the processing another one? Depending on the type of the relationship between an image and a spoken word (e.g. whether the word refers to the image, whether the word and the image's name begin with the same phonemes, or whether they share semantic properties), both the likelihood to look at the image in response to the word as well as the word's speed of recognition in the context of the image will vary. Our first set of studies builds on these findings in order to investigate a potential interaction between the tendency to fixate a related objects in response to a spoken noun and the effect this object has on the processing of the word. The last two sections of this chapter detail theoretical accounts of noun-contingent eye movements as well as their potential effect on visual processing.

1.2 Research paradigms used: the visual world and lexical decision paradigms

After introducing the visual world and lexical decision paradigms, the following two subsections place the experimental parameters of the studies presented in this thesis within the broad range of studies in these two research paradigms.

1.2.1 Visual world paradigm

Noun-contingent eye movements in language comprehension, as described in the introduction, are used in psycholinguistics to study language processing as well as the interaction between visual and linguistic processing. The method used to accomplish this is called the visual-world paradigm, first introduced by Cooper (1974) and revived by Tanenhaus et al. (1995). In a visual-world study, participants are presented with a controlled visual environment to inspect and controlled spoken utterances to comprehend. The listeners' visual attention towards entities in the scene is recorded using an eye-tracker and interpreted as an indication that concepts of the looked-at entities are being considered in the on-line construction of an interpretation of the spoken utterance. Tanenhaus, Magnuson, Dahan, and Chambers (2000) describe a number of advantages of the visual-world paradigm for linguistic research: its high temporal resolution provides insights into the time course of reference resolution even at its intermediate stages, it is non-invasive and thus does not disturb linguistic processing and it can be used in genuine conversational interaction.

Before proceeding, it is worth outlining the standard elements of a visual world experiment, as well as some of the variations that have been exploited.

The visual stimuli are at the core of the paradigm and constitute the environment that makes the activation of concepts accessible via eye-tracking. They sometimes are real objects laid out on a table in front of the participant, but most often are depicted on a computer display. Single objects or persons are arranged in grids (Tanenhaus et al., 1995; Allopenna et al., 1998; Dahan, Magnuson, Tanenhaus, & Hogan, 2001) or in (often simplified) scenes (Altmann & Kamide, 1999). These scenes may be static or they might change over the course of a trial (e.g. Knoeferle and Crocker (2007) use sequences of static clip-art 'scenes' to depict an unfolding event). The screen may also turn blank after an initial scene display (Altmann, 2004). Such blank-screen studies will be introduced in Section 1.5.

The linguistic stimuli consist of spoken single words (Experiment 2 in McMurray, Tanenhaus, Aslin, and Spivey (2003); Creel, Aslin, and Tanenhaus (2008)), entire sentences (Tanenhaus et al., 1995; Altmann & Kamide, 1999; Huettig & McQueen, 2007) or even short stories (Arnold, Eisenband, Brown-Schmidt, and Trueswell (2000)) that are related to the visual stimulus, either directly or indirectly (Yee & Sedivy, 2006; Huettig & McQueen, 2007). They may be semantically or syntactically ambiguous, up to a certain point or even in their entirety (Spivey et al., 2002; Weber, Braun, & Crocker, 2006).

In the majority of visual-world studies, the visual stimuli are presented shortly before the linguistic stimuli, so that the linguistic stimuli can be processed in the context of an already established representation of the current visual environment. Consequently, eye movements can be taken to be driven mainly by the novel, linguistic, stimuli, because the initial processing of the visual stimuli is finished. We use the name 'preview period' for this period of time, during which the visual stimuli are displayed before the onset of the linguistic stimuli. The duration of the preview period varies between visual-world studies and ranges from zero (Experiments 1 and 3 in Kamide, Scheepers, and Altmann (2003); Boland (2005)) to several seconds (Salverda & Altmann, 2005).

The experimenter typically intends for participants to thoroughly process the linguistic stimuli in the context of the visual stimuli, but does not necessarily want the participants to be aware of their eye movements being monitored. Participants are given linguistic tasks (such as comprehension questions: Scheepers, Keller, and Lapata (2008); Richardson and Dale (2005), object manipulation tasks (e.g. to pick up an object and put it somewhere: Tanenhaus et al. (1995); Allopenna et al. (1998), asked to click on or touch an object: Yee and Sedivy (2006)), a passive comprehension task ('look-and-listen': Altmann and Kamide (1999); Knoeferle and Crocker (2006)) or a picture-sentence matching task (Arnold et al., 2000; Staudte & Crocker, 2011).

In all of these settings, eye fixations on the different objects and regions in the display are recorded and analyzed in temporal relation to relevant words in the spoken utterance in order to determine which objects or areas in the display attracted visual attention while certain parts of the utterance were being processed. Looks at an object during the comprehension of the linguistic stimulus can be taken as an indication for different kinds of linguistic and visual processing. Looks to a displayed object during and shortly after its mention are taken to reflect lexical access to its name and its consideration in referent assignment. This linkage is used in studies on noun comprehension (e.g. on lexical access: Allopenna et al. (1998); Dahan et al. (2001)) as well as in studies on parsing and disambiguation (Tanenhaus et al., 1995; Spivey et al., 2002). Looks to display entities shortly before the mention of their names are taken to indicate the anticipation of these objects' mention as a function of the preceding context (Altmann & Kamide, 1999; Kamide, Scheepers, & Altmann, 2003; Kamide, Altmann, & Haywood, 2003; Altmann & Kamide, 2007). As it is our goal to investigate only the speed of lexical access expressed through looks to the referent, other factors that could affect eye movements need to be accounted for as well. Various visual properties (e.g. size, level of detail and position) of an object that determine its visual saliency as well as lexical properties of a noun that determine its likelihood to be recognized (e.g. frequency and length) are controlled in our designs in order to make sure that they do not affect eye movement behavior and confound those effects we set out to investigate.

On the visual-world level of our own experiments, we present participants with a single spoken noun and a grid of objects. Looks to the spoken noun's referent and related objects in a grid of objects are evaluated in order to explore the attraction and effect of those objects' processing while the spoken noun is comprehended.

1.2.2 Lexical decision and priming

A noun-contingent eye movement focuses attention on an object that is associated with, along some dimension, the spoken noun. The processing of one concept being affected by the processing of another concept has been explored by numerous priming and lexical decision studies (Huettig and Altmann (2007) even draw an analogy between noun-contingent eye movements and cross-modal priming). So, in order to determine how the comprehension of a spoken noun is influenced by simultaneous looks at a related object, we exploit an auditory lexical decision task on the spoken noun for the experiments reported in this thesis and turn to previous priming studies to base our hypotheses on.

In an auditory lexical decision task, participants are instructed to respond to a spoken stimulus as quickly as possible, indicating if it is a word or not. This task is often used in priming studies: an additional stimulus, the prime, is presented before or simultaneously with the spoken one. The correctness and, more importantly, the latency of participants' lexical decision response in the context of the prime is taken as an indication of the speed of lexical access for the spoken noun, as it might be influenced by the priming stimulus.[1] As mentioned above, the type of relationship between prime and target is a crucial element of lexical decision studies: it can be identical (prime and target are tokens of the same concept: e.g. Levelt et al. (1991); Dennis Norris and Butterfield (2006)), phonologically related (usually the prime and target's names share their initial phonemes: e.g. Goldinger et al. (1992); McQueen and Huettig (2005)), semantically related (prime and target could be associated or from the same semantic category; e.g. Moss, Ostrin, Tyler, and Marslen-Wilson (1995); McQueen and Huettig (2005)) or unrelated (no phonological, semantic, visual or other relation between the prime and target concepts or tokens). The next section will detail existing research on the different, in part contrary, effects arising from these types of primes.

The temporal characteristics of the stimulus presentation (e.g. stimulus onset asynchrony) are a crucial aspect of priming studies and can affect the observed effects significantly. In many visual-world studies, including the ones presented in this thesis, the visual stimulus (roughly corresponding to the prime) is typically visible for some time before the linguistic stimulus

[1] Some authors (Balota & Robert F. Lorch, 1986; Goldinger, 1996; Seidenberg, Waters, Sanders, & Langer, 1984) assume the existence of post-access processes in lexical decision, for example the attempt to find a relationship between the prime and target after having recognized the target. These processes could lead to an overestimation of beneficial effects e.g. in semantic facilitation.

(corresponding to the target) and also remains on display during the linguistic stimulus' lifetime. This is only true for a subset of priming studies, and we will focus the following review of established priming effects on studies that have these timing characteristics.

In addition to timing, another aspect of the design of priming and lexical decision studies are the prime and target's modalities. Although stimulus modality can be (and is) treated as a factor that has an effect on priming, the sparsity of congruous previous results for each of the types of prime-target relatedness is the reason why this thesis will feature various combinations of prime and target modalities in the following sections about existing findings from priming and lexical decision studies.

By experimentally manipulating which sorts of representations interact with the word recognition process at which points in time the authors of previous priming studies have drawn conclusions on, for example, the organization of the mental lexicon (Balota & Robert F. Lorch, 1986; Holcomb & Anderson, 1993) or details of the comprehension process (Goldinger et al., 1992; Dennis Norris & Butterfield, 2006).

We use an auditory lexical decision task in all of this thesis' experiments in order to assess the speed of comprehension of the spoken word that directs visual attention. As we are also interested in an eye-movement choice made to single out elements of one's visual environment, we display several visual objects, one of which, the critical object, is our lexical decision prime[2] for the spoken word target. This critical visual object can either be related to the spoken word or it can be an unrelated baseline object. Our stimulus-onset asynchrony corresponds to the preview period described in the above section on visual worlds: the visual prime is presented before and during the spoken word.

In the sections below, we present common visual-world and priming effects for different degrees of relatedness between a critical visual object and a spoken noun: reference, phonological relatedness and semantical relatedness.

1.3 Noun-contingent eye movements and priming effects

The studies in this thesis are anchored in both the visual world and priming paradigms, as they investigate how the tendency to fixate related object in response to comprehending a spoken word affects that comprehension process. This section presents previous research about how

[2]One would expect that the simultaneous presentation of additional pictures together with the prime should affect the priming effects otherwise known for only a single picture in the display. To our knowledge this issue has not yet been addressed experimentally. In our own experiments, however, as we do not vary the number of visual objects displayed between conditions in any of our experiments, our conclusions should not be affected by a possible multiple-prime confound. Those priming effects that we do find on multiple-object displays we also find for single-object displays.

hearing a spoken noun affects looks to related objects (NCEs) and how the processing of one concept is affected by the simultaneous processing of a related concept (priming).

1.3.1 Referent objects

In these next two sections, studies on noun-contingent eye movements to referent objects as well as priming from identical primes will be presented.

Effect of linguistic on visual processing: Referent NCEs

The most basic eye movement behavior observed in psycholinguistic visual-world studies is that while comprehending a spoken noun, people tend to shift their gaze towards those objects in their visual environment which are potential referents for this noun, beginning right at word onset. Cooper (1974) conducted the classical study that introduced the visual-world paradigm to psycholinguistic research. His experiment tested and confirmed the hypotheses that when people hear spoken language and see visual elements related to the speech's contents, they will preferentially look at those visual items most closely related to the meaning of the comprehended language, and that these shifts of visual attention occur during or shortly after the pronunciation of the relevant referring expressions. In that study, four short prose stories were played to participants together with visual displays containing nine objects each. There were some instances of direct and indirect reference to the display in each short story. Already in this initial exploratory study, Cooper found evidence for visual target fixations in response to nouns, pronouns, adjectives and verbs, target fixations to words as interpreted in their context, and target fixations beginning before the referring expression was finished, as well as anticipatory eye movements. The more recent study by Tanenhaus et al. (1995) re-establishing Cooper (1974)'s paradigm for modern psycholinguistics found, among other effects, that looks to a spoken noun's referent began approximately 250 msec after the end of the first word that uniquely identifies it. This does not have to be the noun itself, but could be an earlier constituent, like an adjective in Tanenhaus et al. (1995)'s instruction "Touch the starred yellow square". If there were several starred objects but only one starred yellow object, looks to the starred yellow square would be initiated already about 250 msec after the end of "yellow". The issue of incrementality will be addressed again in Section 1.3.2. Since these pioneering experiments, many studies have used referent NCEs in order to investigate various aspects of language comprehension. For instance, Magnuson et al. (2007) examined the partial activation of candidates that are phonologically similar to a comprehended noun, refining computational models of lexical access. Crucially, and similarly to our studies, the visual stimuli in their experiment consist of only the spoken noun's referent and three distractors, so there is no competitor object co-present in the scene (which is common in other studies). Participants were asked to click on objects in the display and there was no preview period before the instruction's onset, so participants had the duration of "Click on the" to inspect the display before fulfilling their task. The time course of participants' likelihood to fixate the referent varied with cohort density, neighborhood density and frequency of the spoken noun. Overall, referent fixation probabilities began to rise from around 200 msec after word onset and peaked around 900 msec after word onset, reaching a maximum fixation probability of .8. Given that the referent is uniquely identifiable only and already at word onset, the probability of fixations on the referent rising at 200 msec means that the first saccades towards it are launched right when the first phoneme is being perceived, as it takes about 200 msec to perform a saccade in visual world settings (see Appendix in Altmann and Kamide (2004) for discussion).

In the study just described, participants were instructed to interact with the named object, essentially forcing them to look at the target object upon comprehension of the spoken noun. Noun-contingent referent looks also occur in settings where the connection between the display and the spoken stimuli is not as explicit and task-driven, as is also the case in our studies: Huettig and Altmann (2005) find 'visual semantic priming' with semantically, but not associatively related items and add category membership to the set of relationships between linguistic and visual stimuli that can drive NCEs. On two out of three conditions of their study, only one object related to the spoken noun was in the visual display, as in Magnuson et al. (2007) and our studies. Before the onset of the critical spoken noun, there was a one-second long preview period and about four seconds' duration of a carrier sentence. Participants did not get an explicit task, they were told simply to 'look and listen'. On the condition on which the display contained a referent and three distractors, Huettig and Altmann (2005) observed a higher probability for participants to fixate the referent than the distractor objects starting from about 150 msec after word onset, with referent inspection probabilities peaking around 600-700 msec after word onset.

Referent NCEs occur consistently in other types of study as well, with competitors co-present in the same display, with referring expressions embedded in sentences, and with scenic visual stimuli.

To sum up this section, previous research shows that looks to objects named by spoken referring expressions occur reliably and right from the spoken word's onset on.

Priming by referents

Priming studies offer some evidence that the comprehension of a spoken noun can benefit from the visual presence of the noun's referent object, but the results are more sparse than one might think. Matsukawa, Snodgrass, and Doniger (2005), in Experiment 1 of their incomplete-picture-identification study, displayed written prime words for 500 msec followed by a 1500-msec break before visual target objects were presented. These visual targets were more likely to be identified correctly when primed by their own names than by unrelated words. More similarly to the present studies' setup, Experiment 3 of Levelt et al. (1991) interspersed an auditory lexical decision task in a picture naming experiment. While participants' prevailing task was to name the displayed picture, they would occasionally hear a spoken stimulus they had to perform a button-press lexical decision task on. This means that participants saw an object drawing they were supposed to name (the prime, in our understanding), and also heard a spoken noun that was the subject of a lexical decision task (the target). On experimental trials, which were a subset of these lexical-decision trials, the spoken stimulus was a noun that was related to the depicted object: it could *name* the object, be phonologically related (with the same onset as the depicted object's name), semantically related or unrelated to it. The stimulus onset asynchrony between the visual object and the spoken stimulus was varied between participants. With one of these SOAs, the spoken stimulus' presentation began on average 673 msec after the picture's onset[3] and the picture remained on the display during the lifetime of the spoken stimulus. At this SOA, lexical decision latencies were shorter when the prime object was the spoken word's referent than when it was an object unrelated tot he spoken word.

In sum, there is evidence for the processing of a concept being beneficial for the subsequent processing of the identical concept in a different modality, specifically for a picture prime and a spoken noun target. We will further explore this effect in the following experiments, where such referent priming is hypothesized to be part of the interaction between visual processing and word comprehension manifest in noun-contingent eye movements.

1.3.2 Phonologically related objects

An object whose name has the same onset as a spoken word is likely to receive looks triggered by the comprehension of that spoken noun, but should have an adverse effect on comprehension compared to referent objects.

Effect of linguistic on visual processing: Onset competitor NCEs

Noun-contingent referent looks, as described above, can be launched at the earliest possible point in time given the available evidence, that is right from word onset. It is an acknowledged behavioral pattern in language comprehension that even partial linguistic input is instantly mapped

[3]The actual SOAs were determined individually for each item and were based on the recognition latencies of the depicted objects, which were determined in pre-tests: the SOA referred to above was exactly 100 msec longer than the the individual picture's recognition latency.

onto internal representations, and this mapping is updated continuously as more linguistic input becomes available. This principle of incremental processing holds not just at the word (cf. Tanenhaus et al. (1995)'s finding in the previous subsection) and sentence levels (cf. studies on anticipatory eye movements like Altmann and Kamide (1999) or Kamide, Scheepers, and Altmann (2003)), but also at the phoneme level: during the comprehension of a spoken noun, all those objects whose names match the currently perceived portion of the spoken noun are looked at equally often, as they are all valid hypotheses about the identity of the referent, until the word's point of disambiguation has been encountered. This has been discovered in lexical access studies that have used the high temporal sensitivity of the visual world methodology to observe the activation of lexical entities during the referent assignment process. In visual-world studies, those objects that are not the referent of, but whose names share initial phonemes with the spoken word are called onset competitors (in our own studies below and in reference to priming studies they will be called phonologically related objects).

The first study to show onset competitor looks was conducted by Allopenna et al. (1998). In their Experiment 1, participants were instructed to pick up an object on a computer screen and put it somewhere else in the display, using a computer mouse. In one condition the display contained the to-be-moved object, an onset competitor, a rhyme competitor and a distractor object. The preview period lasted for about two seconds, then a fixation cross in the center of the display had to be fixated for about 1 second and after that the experimenter spoke the pick-up-and-move instruction. The study showed both referent and onset competitor looks starting from 200 msec after the onset of the referring noun, which means that these fixations must have been launched almost immediately after word onset. Allopenna et al. (1998) established that looks to onset competitors can be launched as soon as the input provides evidence *for* them being the referent. Dahan and Gaskell (2007), on the other hand, explore for how long looks to onset competitors linger beyond the spoken noun's point of disambiguation, which is where evidence starts speaking *against* them being the referent. They show that the probability to fixate an onset competitor begins to drop only 250 msec after the point of disambiguation, which they take to be long after the disambiguation point. However, as 250 msec are just a little longer than the time needed to launch a saccade, the first looks away from the onset competitor would actually have been launched very soon after the point of disambiguation. Onset-competitor NCEs have been shown in numerous visual-world studies on word recognition (e.g., in addition to the above-described ones, Dahan et al. (2001); Huettig and McQueen (2007)) and can be considered a reliable effect. They demonstrate that looks to potential referents of a spoken noun can be launched so early into the noun (during its initial phonemes) that they do not always hit actual referents, which only becomes clear when subsequent mis-matching phonemes of the spoken noun are heard.

Priming by phonologically related objects

The context of a phonologically related object (an object whose name shares its first phonemes with the target's name) has been shown to have a detrimental effect on the processing of a concept when the delay between the onset of the prime and the onset of the target is long enough for participants to be able to process the prime's name.

In an auditory lexical decision task, Goldinger et al. (1992) (Experiment 5) present participants with auditory primes whose first phonemes overlap with the spoken target words', with an inter-stimulus interval of 50 msec. They find interference from those phonologically related primes in comparison to unrelated primes. Two more studies confirm the effect for a setting in which the prime's lifetime lasts beyond the target's onset. Levelt et al. (1991), in the experiment described in Section 1.3.1, find responses in an auditory lexical decision task to be slower when the spoken word is primed by the image of an object that is phonologically related to the spoken stimulus than when it is primed by an unrelated object image. Similarly, McQueen and Huettig (2005) present participants with single prime pictures that can be phonologically related to auditory lexical-decision target words. The prime picture is displayed from 2400 msec before the onset of the target word, and stays on the display until the lexical decision has been made. This study produced slower lexical decisions with phonologically related primes than with unrelated primes.

In sum, there is evidence that phonologically related depicted objects (onset competitors), act as primes that can interfere with the spoken noun's comprehension. As such objects are also likely targets of noun-contingent eye movements (cf. the previous section), they constitute interesting stimuli for studying the interaction of NCEs and noun comprehension: They are often looked at upon noun comprehension, but should at the same time affect comprehension negatively. The first experiments of this thesis further explore this seeming conflict in order to find out more about the motivation behind noun-contingent eye movements.

1.3.3 Semantically related objects

Even though not visually similar to a spoken noun's referent, an object whose concept is semantically related to the the spoken noun's attracts looks upon noun comprehension and may have a beneficial effect on the comprehension of the spoken noun.

Effect of linguistic on visual processing: Semantically related competitor NCEs

While looks to referents and onset competitors are evidently motivated by the search for the spoken word's referent, there are also less readily explained noun-contingent eye movements: looks to objects semantically related to the spoken word. Initially observed already in Cooper (1974)'s study, looks to semantic associates upon comprehending a noun have been shown by Yee and Sedivy (2001) in a visual-world study and by Moores, Laiti, and Chelazzi (2003) in the visual search for a written word's referent. Similar to the studies in this thesis, Huettig and Altmann (2005) used objects from the same semantic category as the spoken noun as semantic competitors in a visual-world experiment. They displayed either a referent and three distractor objects, a semantic competitor and three distractors, or the referent, the semantic competitor and two distractors. Participants were to listen to a spoken sentence without an additional task. This sentence was played after a one-second preview period and lasted on average four seconds before the mention of the referent's name. Crucially for our studies, Huettig and Altmann (2005) found that participants looked at the semantic competitor more than at the distractors when the referent was not co-present in the display. When the referent was displayed, the semantic competitor still received more looks than the distractors, but less than the referent. Also, when the semantic competitor was depicted with distractors alone, the probability for the semantic competitor to be looked at rose from 250 msec after the onset of the critical noun, peaked around 600 - 700 msec after word onset and did not drop within the 850 msec after word onset that were plotted by Huettig and Altmann (2005). Noun-contingent eye movements to semantically related objects were replicated by Yee and Sedivy (2006) using slightly different experimental design: In Experiment 1 of their study, the spoken noun's referent, a semantically related object, an unrelated control object and a distractor were displayed simultaneously with the spoken instruction to touch the referent object. One could hypothesize that the instruction to touch the referent might make looks to a semantic competitor interfere with the task performance and that semantic competitor looks might thus be less likely to occur with this task than in Huettig and Altmann (2005)'s study, where there was no task that these looks could have interfered with. Still, Yee and Sedivy (2006) found more looks to the semantic competitor than to the control object starting from about 200 msec after word onset.

These looks to objects semantically related to a spoken noun have been interpreted as a by-product of the lexical access process, where not only those concepts whose names overlap with the comprehended noun are activated, but activation spreads to concepts related to the spoken word's concept and triggers looks to their images as well as to referent objects (Huettig & Altmann, 2004; Yee & Sedivy, 2006; Huettig & Altmann, 2007).

Priming by semantically related objects

There is an abundance of evidence for a beneficial effect of the processing of a semantically related object on the processing of another concept, both for when the prime's duration does and does not overlap with the presentation of the target. An auditory lexical decision experiment with auditory prime and target words by Moss et al. (1995) showed facilitation on trials with semantically related primes compared to trials with unrelated primes, at inter-stimulus intervals of 200 and 1000 msec. A study by Koivisto and Revonsuo (2000) shows faster lexical decisions for written targets preceded by 500 msec by semantically related pictorial primes with a duration of 150 msec. Similarly, but with a visual context (prime) that temporally overlaps with the target, McQueen and Huettig (2005) (described in Section 1.3.2) present visual primes as primes 2400 msec before spoken target words, with the picture remaining on the display throughout the spoken word. Lexical decision latencies were found to be slightly shorter on trials with semantically related primes than on trials with unrelated primes. Also, Holcomb and Anderson (1993) could show semantic facilitation in an auditory lexical decision task when short-lived (duration of 200 msec) written prime nouns were presented simultaneously with their auditory targets. Lastly, a lexical decision experiment with auditory primes, auditory targets and a 200-msec stimulus-onset asynchrony produced a semantic facilitation effect in Anderson and Holcomb (1995).

These findings offer at least preliminary evidence that in a visual-world setting, the presence of a semantically related object could be a source of a benefit for the processing of a spoken noun.

1.4 Mechanisms behind noun-contingent referent looks

The previous section reviewed visual world research and illustrated three types of noun-contingent eye movements: referent looks and looks to phonologically or semantically related objects. It also reviewed a number of priming and lexical decision studies in order to show that the processing of another concept simultaneously with or slightly before the processing of a spoken noun can affect that noun's recognition. The notion of priming as a process that is relevant for explaining the motivation behind noun-contingent eye movements will be taken up again in the Section 1.5. This section will review how noun-contingent referent looks, which are the focus of this thesis, have been integrated in theoretical frameworks by other researchers.

First, we will recall the phenomenon that we set out to explain: During a preview period that can last from a few hundred milliseconds to a few seconds, people establish the identity of the objects in their visual environment. Then, a spoken word that names one of the objects in the environment is perceived, and (partial) comprehension of the name triggers an eye movement to this object. Researchers studying language comprehension in the visual-world paradigm predominantly adopt a combination of the following theories in order to explain such noun-contingent eye movement behavior: External memory or situated processing accounts are used to explain how cognitive processes can exploit, in addition to internal representations, information in the environment. Spatial indices are described as part of internal representations and as the mechanism that makes one's visual environment accessible to cognitive processing. Finally, aspects of embodiment theories characterize the nature of memory representations to include sensory and motoric components, providing a context for the spatial indexing mechanisms. In the following paragraphs, we will describe how these theories can be combined to form a cohesive account of noun-contingent eye movements.

External memory and situated processing O'Regan (1992), in presenting a novel theory of visual processing, introduces the notion of the external environment as part of the human memory: O'Regan states that, instead of creating and maintaining a detailed internal representation of one's visual environment, people sample the environment directly via fixations when and as required by the current task. Thus, the visual environment is used as an external memory resource. Ballard, Hayhoe, Pook, and Rao (1997) adopt a similar principle in a computationally motivated account of cognitive processing in general: according to them, information can be acquired from the environment on-line whenever necessary, because storing it in working memory can be more expensive. Ballard et al. (1997) base this argument on considerations about computational efficiency and support it with data observed in a block-copying task.

An externalist theoretical account of speech-contingent eye movements is given in Spivey et al. (2004). They stress the mind's embodied and embedded nature: the mind is part of a body and dependent on, as well as shaped by, the body's sensoric and motor systems; it is also part of an environment that participates in shaping cognition. Information can be accessed perceptually from the environment as quickly as cognitively from memory, so that the external world can function as a part of the human memory. Visual perception is a major instance of the mind using the world as its own memory, where the visual environment is probed actively and continuously through fixations.

The term situated language processing is used predominantly in studies about spoken sentence comprehension in visual world settings. It denotes the co-operation of several different processing domains in the creation of an utterance interpretation that integrates knowledge about the current situation with the linguistic input, and conceptual as well as episodic memory (Altmann & Kamide, 2009). Within this framework, external resources, i.e. parts of the visual scene, are accessed via eye movements when the spoken utterance makes them relevant by referring to them (whether directly or indirectly). This is because the system seeks to integrate situational knowledge into the interpretation of the utterance and habitually accesses this type of information from the external environment when necessary (Huettig & Altmann, 2004).

Spatial pointers as the mechanism behind noun-contingent eye movements In external memory and situated language processing theories, the cognitive construct through which external visual information can be accessed when necessary is a spatial index or spatial pointer. The notion of the spatial index was carved out by visual attention researchers like Daniel Kahneman, Anne Treisman and Zenon Pylyshyn, who created the basis for it in their object-based theories of visual attention. In a study of spatial language and visual processing, Logan (1995) defines a spatial index as a mental symbol attached to a location or object in the perceptual representation of one's visual environment, i.e. a symbolic address. Spivey et al. (2004) spell out the definition of spatial indexing in greater detail: In their account of externalist processing, a spatial pointer is composed of the address of an object's location in an environment and a label. This label is a sparse internal description of the object at the pointer's address, not being any more informative than necessary in order to identify objects in terms of the current task demands. Task planning operates on these labels and when it requires information on or interaction with an object in the environment, it activates its label and thus the address associated to it, leading to a look towards the address where relevant external information can be taken in. In these terms, the comprehension of a spoken noun would activate the label of the referent's spatial index, as the entity the comprehended utterance refers to is relevant in the task of comprehension, and thus cause an eye movement to the indexed entity, the referent. The spatial index is placed in a language comprehension context by Altmann and Kamide (2009). Incorporating elements of the embodiment theory, they see language comprehension as a process that produces a simulation of the described event by integrating linguistic and all kinds of other available kinds of information, like visual information, contextual information or world knowledge. During the preview in a visual-world setup, episodic memories of the perceived objects would be created, which contain spatial pointers to these objects. When an utterance is then comprehended and a noun is encountered, the simulation of the described event will activate the episodic memory of the named object and with it the spatial pointer that causes a new eye movement towards the noun's referent.

1.5 The potential of NCEs to benefit noun comprehension

While it is conceivable that noun-contingent referent looks are merely a consequence of visual and linguistic processing being intertwined as described in the previous section, this thesis explores the idea that NCEs might have a beneficial effect on noun comprehension and thus a purpose. The basic mechanism that we hypothesize to lie at the heart of such a benefit is this: The noun comprehension process starts out with the activation of a number of referent candidates (McQueen, Dahan, & Cutler, 2004), which can all trigger noun-contingent eye movements (if there is a visible referent, some of these eye movements will be directed towards the referent, some will potentially be directed towards phonologically related objects - up to the noun's point of disambiguation, all these objects are likely referent candidates; cf. Section 1.3.2). These eye movements are launched from the earliest point in time at which the available evidence singles them out and thus before the recognition of the spoken word is completed. If the referent is looked at, the boost of activation to its representation can feed back into the word recognition process and raise the activation of the referent candidate, possibly enabling faster word recognition than without noun-contingent eye movements. Out of this whole process, this thesis (in Experiments 5 to 7) deals with the ways in which the information provided by noun-contingent could be beneficial for noun comprehension. In the paragraphs below, we will discuss how the external and internal information accessed by NCEs could facilitate the recognition of a spoken noun.

Visual instance of the referent Most obviously, a noun-contingent referent look will provide one with the visual image of the spoken noun's referent. There are a number of means by which this visual instance of the referent could be beneficial for the word recognition process, as described in the following paragraphs.

Referent perception benefit: As described in the above section on priming by referents (Section 1.3.1), there is evidence that the processing of a referent picture concurrently with the comprehension of a noun can be beneficial for the processing of that noun. In a continuous word recognition process in which several hypotheses about the word's identity are entertained in parallel until the system settles on one (McQueen et al., 2004), the boost that the word's concept could receive by its visual instance would likely speed up its selection and thus the word's recognition.

Preservation of priming: The above section on referent priming also cites a study demonstrating a beneficial effect on noun comprehension from the processing of an identical prime *before* the noun's onset. It is thus conceivable that noun comprehension is facilitated by the fact that one has seen the noun's referent during the preview period. It could then be irrelevant where people look during comprehension, because the priming effect could be in effect from the preview period alone. Alternatively, referent looks during the comprehension of the noun could be a strategy that *sustains* this priming effect (i.e. it would be targeted at an object that does not inhibit the priming effect), but not have a beneficial effect themselves.

Avoidance of interference: Griffin (2004), in a paper that thoroughly examines several hypotheses for a potential benefit of name-related gazes in language production, also introduces the so-called interference-avoidance hypothesis, which asserts that name-related gazes are made in order to avoid interference from fixating unrelated objects. The evidence presented by Griffin (2004) implies that the avoidance of interference is a factor driving speech-related eye movements at the stage of message planning in language production, especially in complex utterances. We will explore, in theory and in Experiment 5, whether a similar hypothesis holds for speech comprehension: are noun-contingent referent looks beneficial for the comprehension of the noun triggering them because they avoid the interference that would arise if one were to look at unrelated objects instead? On a general processing level, Arthur Glenberg and Robertson (1998) show that the recall of information benefits from the aversion of gaze from engaging visual stimuli when one tries to recall complex memories. They conclude that the system can choose between information intake from internal and external resources and as both types of probing take effort, it can be beneficial for one type of retrieval to suppress the other. Some priming experiments have examined the effect of unrelated and absent primes on noun processing and might give us an indication of whether seeing an unrelated object might have a detrimental effect on noun comprehension: Two early written lexical decision studies with, among others, unrelated and non-word (several 'X's) primes (Neely, 1976, 1977) showed slower lexical decision latencies when primes were unrelated words than when they were non-words, at various negative stimulus-onset asynchronies (i.e. primes presented before targets). Assuming that a non-sensical entity (an 'XXXX') has no capacity to aid the comprehension of the target word, this is attributed to interference from the unrelated primes. Also, Experiment 3 of Levelt et al. (1991), described earlier, provides evidence that auditory words without the context of picture primes were processed faster than auditory words presented concurrently with a visual object - both for visual objects that are related as well as unrelated to the spoken word.

Experiment 5 of this thesis will test the three hypotheses described in this section on the ways in which referent's visual instance, taken in during a noun-contingent referent look, could be beneficial for word comprehension.

Location of the referent and associated information There is reliable evidence for noun-contingent eye movements being launched even when the referent of the spoken noun is no longer present during noun comprehension, in a so-called blank-screen paradigm (Altmann, 2004; Altmann & Kamide, 2004). In Altmann (2004), a simple scene (containing two persons and two objects) was displayed for five seconds, then the display went blank at the same time as participants heard sentences like "The man will eat the cake.". Saccades towards the cake were launched reliably before and during the word "cake" (the verb "eat" permitted anticipation of its direct object in the context of the depicted entities). These blank-screen noun-contingent eye movements could be merely an automatic consequence of the activation of the object's episodic trace (Altmann, 2004): As one's visual environment in 'real life' will not usually go blank from one moment to another, re-fixating the location at which a referent object had been perceived earlier will often provide visual information that matches the spoken noun. It is conceivable that this behavior is so well-learned that it automatically perseveres even in situations where NCEs will not actually give access to any visual referent representation (i.e. no beneficial visual information) at all, such as in a blank-screen setting. Alternatively, even NCEs on a blank screen could potentially provide information that is beneficial for word comprehension. This section discusses what kinds of information this could be.

A similar question has been discussed in detail for looks to blank locations while recalling facts associated to these locations, in the so-called 'Hollywood Squares' paradigm and related studies (Richardson & Spivey, 2000; Richardson & Kirkham, 2004; Hoover & Richardson, 2008). During the initial part of such an experiment, participants listen to spoken statements while their visual attention is focused on certain persons (who might be delivering the statements) or regions in the display. When having to recall a fact during the ensuing question session, participants are likely to re-fixate the display region which they were fixating while listening to that fact. These looks during fact recall happen whether the display still contains the information shown there during preview, which bears no direct relation to the recalled fact's contents, or whether the display has gone blank in the meantime. None of these studies, however, find a relationship between participants' likelihood to re-fixate those display locations they had previously associated with a fact and the precision or speed of that fact's recall. Richardson, Altmann, Spivey, and Hoover (2009) summarize their own previous accounts of these results as follows: In situations in which one accesses short-term fact memory, one's visual environment is usually roughly the same at the time a fact is recalled as at the time when the fact was memorized. Those entities in the visual environment that were looked at during the encoding of a fact will, upon recall, constitute cues that can aid the recall of the fact via associative links. In the case of the display going blank between encoding and recall, however, any cues that could be accessed in order to aid fact recall are not present anymore, so no benefit can arise from looks to these cues' previous locations. Then why do the Hollywood Squares studies still observe the blank-screen eye movements during fact recall that they did observe? Richardson et al. (2009) refer back to Altmann (2004) and, in an analogy to blank-screen NCEs, assume that blank-screen looks are not made because people do not notice that the screen has gone blank (as was still stated in Spivey et al. (2004)). Instead, blank-screen looks to the location a fact is associated to are a generalization from a behavior that is usually beneficial, i.e. that would provide cues beneficial for the retrieval of the fact from memory if the visual environments had remained largely unchanged since the encoding of the fact. In sum, in fact recall the access of the mere location a fact was associated to earlier does not provide any benefit.

The question about a beneficial effect of blank-screen noun-contingent eye movements may still be asked and addressed empirically, however. There are different processes at play for these eye movements than in the Hollywood Squares setup. Here, new perceptual (linguistic) information is to be recognized. Looks to the location that matching visual information has been perceived and encoded at might make accessible some sort of information that is of use in this recognition process. What could this information be? According to a 'strict' spatial-indexing account, a blank-screen eye movement should be futile, as it does not give access to any external visual information.[4] In analogy to the fact-recall account described above, however, blank-screen noun-contingent referent looks might provide easier access to the internal representation of the referent perceived during the preview. This memory item, in turn, matches the spoken word and could thus already provide a benefit for comprehension. It is thus still conceivable that blank-screen NCEs could be beneficial for word comprehension even in spite of Hollywood-Squares evidence speaking against a similar benefit in fact recall. With NCEs, the memory item associated to and accessed through the looked-at location could directly benefit comprehension. In fact recall, however, the memory item would only act as an intermediary that aids the access to the to-be-recalled factual information. It may be that only actually visually perceiving a representation of that memory item enables it to act as a cue for the recall of a fact, but that its activation is sufficient to facilitate the recognition of a strongly related single spoken word. Support for this view, however, will require direct evidence that looks to a blank location could actually facilitate activation of visual information previously perceived there. Ferreira, Apel, and Henderson (2008) endorse this claim, referring to studies by Andrew Hollingworth in its support: Andrew Hollingworth (Hollingworth, 2006, 2007) conducted a number of studies which show

[4] This does not mean though that a strict spatial-indexing account could not explain why NCEs are made - as explained above, they could be a generalization from an otherwise (when entities are actually visible) useful behavior.

an improved memory for visual properties of displayed objects when they were perceived in the same location (sometimes as determined by visual context) during recall as when they were memorized. Thus, we can conclude that it is reasonable to test experimentally whether the mere access to a previously seen referent's location in the visual environment can be beneficial for the processing of a spoken noun. We will take up this challenge in Experiments 6 and 7.

1.6 Thesis outline

The behavioral phenomenon at the focus of the research presented here, noun-contingent eye movements, is illustrated in Section 1.3 above. Beginning at the earliest stages of spoken-noun comprehension, people's gaze singles out those objects in their visual environment that stand in a relation to the noun. Visual-world studies have shown that the comprehension of a spoken noun triggers, immediately after word onset, looks towards potential referent objects, towards objects whose names have the same onset as the spoken noun as well as towards objects that have a semantic relation to the spoken noun. This thesis investigates whether there might be a purpose behind this behavior: do noun-contingent eye movements have an impact, potentially a positive one, on the comprehension of a spoken noun? If so, precisely which mechanisms could underlie such an effect from eye movements on language processing?

The theoretical frameworks of situated language processing and external memory can motivate the existence of noun-contingent eye movements by ascribing to them the function to gather relevant external information immediately whenever it is relevant to the mind's current objective. However, these theories ascribe functionality to eye movements on rather high cognitive levels (sentence comprehension, complex problem solving). From a lexical-access perspective, the image of a noun's referent could speed up that noun's comprehension on a much lower level, by boosting its activation and thus enabling an earlier word recognition. This idea motivates our search for a beneficial effect of noun-contingent referent looks on noun comprehension.

The tendency to fixate related objects while perceiving a spoken noun generates a situation in which two related concepts are processed simultaneously. Such situations are created routinely in priming and lexical decision studies, which often find that the activation of one concept has an influence on the processing of a second one. The lexical decision task provides a means to assess the speed of recognition of spoken nouns as affected by a processing of another concept. We utilize this method and combine it with a visual-world setting, creating a novel experimental paradigm that enables us to simultaneously study noun-contingent eye movements within an array of objects and the speed of comprehension of the noun that can trigger these eye movements. In this experimental setting, participants freely inspect an array of objects during a preview period and hear a spoken noun which may be related to one of the objects in the display. During that noun's lifetime the display may remain static and identical to the preview period, or it may provide visual cues that direct participants' visual attention to displayed entities, permitting conclusions about the effect of the displayed entities or looks to these entities on the speed of comprehension of the spoken noun.

In Chapter 2, we use different kinds of related objects as targets of NCEs in order to gain an impression of when eye movements to these objects are beneficial or disadvantageous for noun comprehension and how this correlates to the likelihood of actually executing NCEs to such objects. Experiments 1 and 2 presented in this chapter are pilot studies that observe the co-occurrence of voluntary noun-contingent eye movements to objects related to a spoken noun and the effect that perceiving these related objects has the noun's recognition speed. By manipulating the deployment of visual attention during noun comprehension via a highlighting cue, Experiment 3 investigates how looks to related objects *affect* the comprehension of a spoken noun. Experiment 4, a simple priming study without a visual-world aspect, is constructed to validate the previous experiments' results. This first experimental chapter most crucially establishes that noun-contingent looks to a spoken noun's referent are indeed beneficial for the comprehension of that spoken noun.

The approach of priming and lexical decision studies not only provides methodological features to our studies but also a set of hypotheses about the way in which looks to a referent could create a benefit for noun comprehension. Priming effects can be distinguished by the timing with which prime and target stimuli are processed relative to one another and also by whether one prime is beneficial or another one is interfering for the processing of a target. In Chapter 3, Experiment 5 investigates the nature and timing of the priming effect through which the referent image perceived during noun-contingent referent looks can speed up noun recognition. By manipulating whether participants are presented with the spoken noun's referent at all and whether they look at that referent, an unrelated object or empty space while hearing the noun, three hypotheses can be tested: Firstly, it can be hypothesized that the noun-comprehension benefit caused by noun-contingent referent looks stems from the referent's encoding during the preview period, and referent looks during the spoken noun are only there to preserve this benefit. Secondly, looks to the referent could be beneficial only by preventing interference that would arise if other, unrelated objects were fixated during the comprehension of the spoken noun. Our results provide no evidence supporting any of these hypotheses but speak for the noun comprehension benefit arising only from noun-contingent referent looks *after* the onset of the spoken noun and from a positive effect of referent looks themselves.

Lastly, it can be asked within the spatial indexing framework whether the activation of a spatial index, which is proposed as a mechanism underlying noun-contingent eye movements, not only mediates the access to external visual information but also cues the access to *internal* information previously encoded at that location. Questions about a beneficial function of such internal information have been asked for fact-recall tasks before, but not in the context of spoken-noun comprehension. Testing the hypothesis that access to the location of a previously seen referent could activate the associated internal representation of that referent, which in turn could benefit noun comprehension, Chapter 4 presents two experiments that manipulate the contents presented during noun comprehension at a previously seen referent's location. It can be shown that the looks to a previously seen referent's location do activate internal information in the form of an expectation to encounter the spoken noun's referent, which creates an interference effect when the information actually encountered in this location is unrelated to the spoken noun.

The results of the experiments presented in Chapters 2 to 4 are discussed and put into perspective in Chapter 5, where we conclude that noun-contingent referent looks are beneficial for the comprehension of spoken words; primarily by providing visual input that matches the spoken word's meaning and thus boosts the activation of the appropriate concept. Motivated by Experiment 7's results, we claim that the spatial indexing mechanism, which directs noun-contingent eye movement, also serves as a filter that mediates the influence of perceived visual objects on noun-comprehension.

2 Related objects and spoken word comprehension in visual worlds

The two experiments in this chapter take the first step towards our goal of investigating the interaction between NCEs and noun comprehension. Is it possible to observe a relation between the potentially beneficial or detrimental effects of certain displayed objects on word comprehension and participants' readiness to look at them while trying to understand a spoken word? In order to explore this issue, our experiments combine a visual-world setting with a lexical decision task. We can thus simultaneously and unobtrusively assess the degree to which the comprehension of a spoken noun directs visual attention to objects in the visual environment as well as the speed with which this spoken noun is understood (cf. Section 1.2). As outlined above (cf. Section 1.3), the comprehension of a spoken noun has been shown to trigger looks to its referent as well as to phonologically or semantically related objects. It can be expected, based on the findings of previous visual world studies, that participants begin to shift their gaze towards an object during the lifetime of a spoken word when that object is the spoken word's referent (Tanenhaus et al., 1995; Magnuson et al., 2007) or a phonologically (Allopenna et al., 1998; Dahan & Gaskell, 2007) or semantically related object (McQueen & Huettig, 2005; Yee & Sedivy, 2006). Also, perceiving depictions of objects related to a spoken word should affect the speed of comprehension for that word. There are priming studies that predict beneficial effects of both referent (Levelt et al., 1991) and semantically related objects (McQueen & Huettig, 2005) and detrimental effects of phonologically related objects (Levelt et al., 1991; McQueen & Huettig, 2005) on lexical decision times (cf. Section 1.3). So, these three types of objects all attract visual attention during noun comprehension, but seeing them should have opposite effects on comprehension (beneficial for referent and semantically related objects, interfering for phonologically related objects). This conflict lies at the core of Experiment 1's design, where we manipulated the type of related objects available in a display while participants had to make a lexical decision on a spoken noun.

As congruent evidence for priming in a setup like Experiment 1's is rather sparse (especially for referent and phonologically related primes, cf. Section 1.3) and a lexical decision task has never been studied in combination with a visual-world setup before, the first two studies of this thesis are treated as pilot studies. They are intended to provide initial insights into how the comprehension of a spoken noun and the probability to look at objects related to this noun interact.

2.1 First pilot study: Experiment 1

In contrast to previous visual world studies, Experiment 1's task explicitly stresses the importance of effectively and quickly comprehending the spoken word by virtue of the additional lexical decision task. The first question the experiment addresses is whether, under these circumstances, the likelihood to look at a related object in response to the spoken word varies depending on whether that object tends to have a beneficial or interfering effect on the lexical decision. Also, in contrast to previous priming studies, our participants could choose whether they would look at the object related to the spoken word during word comprehension. It is another open question whether a related object's presence could affect lexical decision latencies in such a setting.

In Experiment 1, participants looked at a computer display containing four objects and after a while heard a spoken word or non-word. Their task was to decide whether what they had heard was an actual German word or not, by saying 'yes' or 'no'. On experimental trials, the spoken stimuli were always words, and one of the displayed objects could be related to the spoken word (cf. Figures 2.1 and 2.2). In this and the following experiments, we will use the term *critical object* to denote this object. The critical object is the object whose effect on both visual attention and speed of lexical decision is of central interest, and whose identity (and thus relation to the spoken word) was manipulated in the experiment. In Experiment 1, the critical object could either be the spoken word's referent, an object whose name began with the same phonemes as the spoken word, an object that was semantically related to the spoken word or an entirely unrelated object.

An issue that is important to consider in priming studies is the question of what kind of a prime constitutes an adequate baseline condition. This is the neutral condition in relation to which benefits or interference from processing critical types of primes have to be compared. In our own experiments, the auditory lexical decision baseline is measured in the context of a visual object unrelated to the spoken word. One could object that this baseline is not strictly neutral, as the unrelated object might have an effect on the processing of the spoken word itself (a question that we address specifically in Experiment 5). However, we defend the choice of this baseline condition on the following grounds: Firstly, the behavior that we want to study is the choice between entities in the real world that people can look at while comprehending spoken

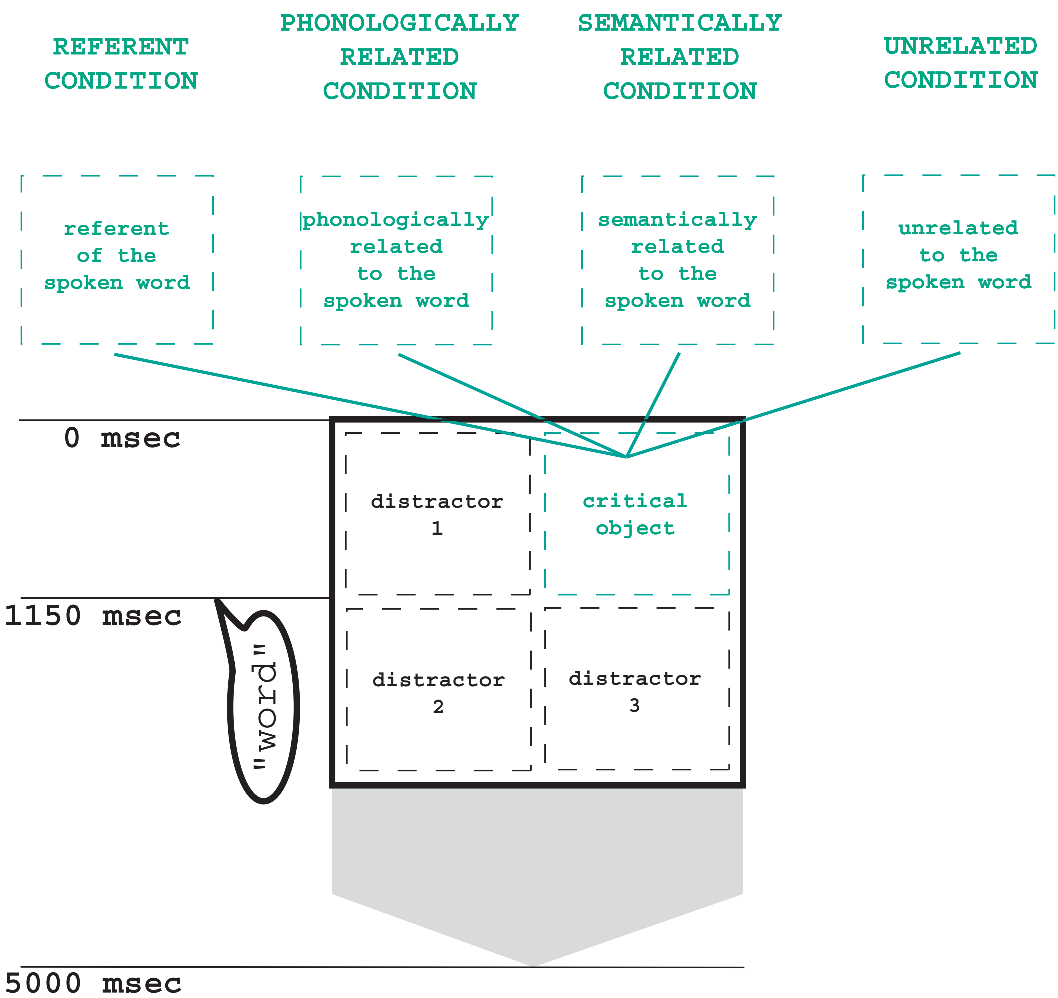

Figure 2.1: A sample trial of Experiment 1. The display, containing the critical object (the spoken word's referent / a phonologically related object / a semantically related object / an unrelated object) and three distractors, was shown with a preview of 1150 msec. Then, the spoken word was played while the display remained unchanged. Participants gave their lexical decision response orally. Every trial had an overall duration of five seconds.

language - that is, we assume that, in a 'visual-world situation', people don't close their eyes to avert gaze and that they will only have meaningful entities to choose between as the goal of their visual attention. They are thus only left to gaze at entities that are either related or unrelated to the perceived utterance, and those unrelated entities are a sensible baseline for evaluating the processing of related entities. Also, following Jonides and Mack (1984), unrelated primes should provide better baseline conditions than meaningless or no primes, because neutral primes should ideally be as similar in appearance, occurrence and informativeness as related primes and differ from them just in terms of their relatedness to the target.

The three objects displayed together with the critical object in Experiment 1 were all unrelated distractor objects. The term *distractor* will henceforth be used for those objects that form a baseline for our visual attention analyses, i.e. the object or objects which remained the same across conditions and could be assumed, due to their unrelatedness to the spoken word and the critical object, not to attract varying amounts of visual attention themselves on the different conditions.

As mentioned above, Experiment 1 is a pilot study and thus driven not by concrete hypotheses, but by an open-ended question. This question is how the influence that the visual perception and recognition of a related object have on the comprehension of a spoken noun perseveres when participants have the freedom to choose whiether they will look at this related object while comprehending the spoken noun.

2.1.1 Method

Participants

Thirty German native speakers from the Saarland University community with normal or corrected-to-normal vision took part in the experiment in exchange for a payment of five Euros. They were largely unfamiliar with the concept of eye tracking. Six participants' data were excluded from the analyses because their responses had not been not recorded properly, because of disturbances during the experiment or because it became clear during the experiment that they almost never looked at the displayed objects. Thus, 24 participants contributed to the analyses.

Materials

Stimuli for each experimental item were assembled from a set of seven inanimate objects (cf. sample stimulus in Figure 2.2): the referent object, an object whose name was phonologically related to the spoken word, a semantically related object, an unrelated object (these were the four critical objects) and three unrelated distractors.

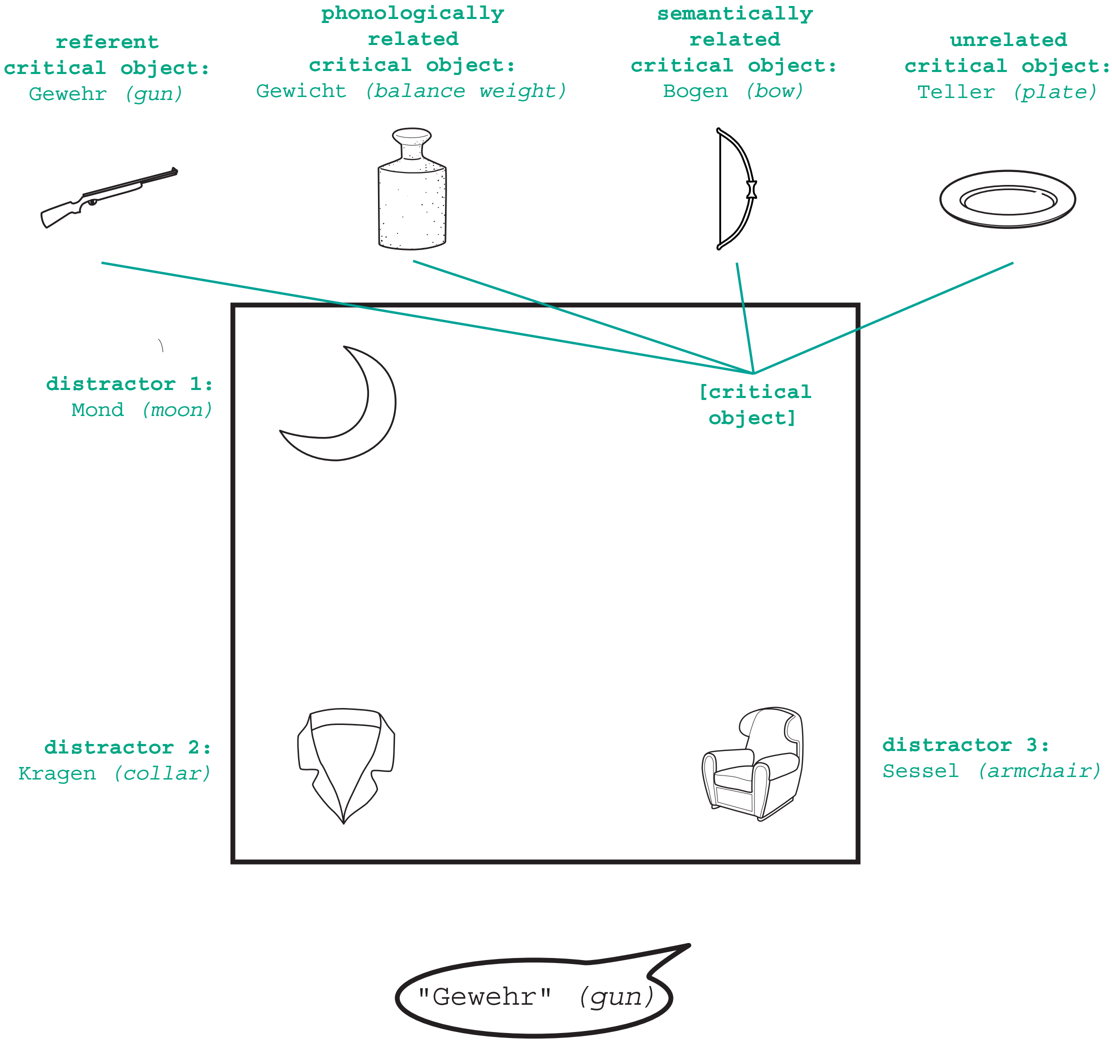

Figure 2.2: The materials for one of Experiment 1's items. Participants were presented with a display containing four inanimate objects: the critical object and three distractors. The critical object's identity was determined by the condition: it could either be the spoken noun's referent, a phonologically related object, a semantically related object or an unrelated object.

An item's spoken word was always the referent object's name, for example 'Gewehr' (gun). Only one of the four critical objects would appear in the display together with the three distractors on any given condition. The relations between the referent object and each of the three other critical objects, and between each of the critical objects and the distractors were controlled as follows:

The phonologically related critical object's name had the same onset as the spoken word, sharing between their first two and their first four (on average their first 2.6) phonemes.[5] In addition, the phonologically related object's name had the same stress pattern as the spoken word. The phonologically related object in our example item, belonging to the spoken word 'Gewehr', was 'Gewicht' (balance weight). The other critical objects as well as the distractors had onsets different from the spoken word, and the distractors' onsets also differed among each other. Naming pretests were conducted to ensure that all depicted objects would be named and identified by participants as intended (see below for details). As the phonologically related critical objects' relatedness to the spoken word hinged on them being named a certain way, especially strict criteria were applied to these objects in the item selection process, based on the results of these naming pretests (cf. below).

The semantically related object was selected intuitively to belong to the same semantic category as the spoken word, meaning that they could be found in the same situations or be used for similar purposes. In our example item, the semantically related critical object was 'Bogen' (bow, which is in the same semantic category as the spoken word 'gun') All other objects were selected intuitively to be semantically unrelated among each other. Pretests were conducted to establish that semantically related critical objects were indeed recognized as semantically related to the spoken word, and that the other objects were unrelated from each other (cf. below).

The unrelated critical object in our example item was 'Teller' (plate), and the three distractor objects were 'Mond' (moon), 'Kragen' (collar) and 'Sessel' (armchair). As mentioned above, the unrelated critical object was selected to be semantically unrelated to the spoken word (as verified in pretests) and to the distractors, which were in turn selected intuitively to be semantically unrelated to all objects of an item.

In addition to name onset and semantic relatedness, the objects used in an item were controlled for name length, name frequency and visual object properties. We did not control for neighborhood properties or uniqueness points of the object names.[6]

[5] If the last shared phoneme was a vowel, the following consonants of the two words had a similar place of articulation, so that the last shared vowel would sound similar for both words.

[6] Since we were interested only in how the speed of lexical decisions on the same word was affected by manipulations of the visual display, these factors, though they are known to affect lexical decision latencies (Goldinger, 1996) were not relevant for our studies.

Name length was measured as the number of syllables and was typically the same for all critical objects of an item. The only exceptions were one item with the phonologically related object's name and one with the unrelated object's name being longer than the referent's name by one syllable, and three items with the semantically related object's name being shorter than the referent's name by one syllable. It was necessary to allow for such minor deviations due to the general difficulty in finding stimuli that satisfied all crucial constraints. The distractor objects' names differed from the referent's name length mostly by one and occasionally by two syllables, making each of the distractors' names longer than the referent's name by on average between 0.2 and 0.3 syllables.

Name frequency was computed based on the frequency classes in the German on-line dictionary Leipziger Wortschatzlexikon "Deutscher Wortschatz" (2007).[7] Due to the impossibility of exactly matching the frequencies of all critical object names, the following criteria were adopted: phonologically related and unrelated objects' names could be more frequent than the referent's name by up to four frequency classes, yielding an average difference to the referent's name of 1.7 frequency classes for the phonologically related object and 1.3 for the unrelated object[8]. The semantically related object's name could be up to three frequency classes more or less frequent than the referent's name, making it on average more frequent than the referent's name by 0.4 frequency classes. Distractors' names could also be more or less frequent than the referent's name by up to three frequency classes and on average differed from it by 0.3 frequency classes.

The pictures used in this study were black-and-white line-drawings. The non-referent critical objects of an item were visually dissimilar from the referent object and all distractor objects were visually dissimilar from the critical objects as well as from each other. The four objects displayed on the screen simultaneously were arranged in a 2 x 2 grid (cf. figure 2.1). Each object filled a square of 188 x 188 px, which stretched across 5.8 °of visual angle. The position of the critical object within the square grid was fixed for each item, but across items the critical object occupied each of the four positions about equally often.[9]

[7] A word's (w) frequency class (fc) in this dictionary is defined as fc(w) = floor($\log_2$(number of occurrences of 'der'/number of occurrences of w), where 'der' (the) is the most frequent German word.

[8] Although (Miozzo & Caramazza, 2003) find larger semantic interference from low-frequency than from high-frequency distractors in a picture-word interference task and our goal in designing our stimuli was to maximize the impact of our critical objects on lexical decision, phonologically related and unrelated critical objects were selected to be *more* frequent than referent critical objects. The motivation behind this was that objects with more frequent names are more likely to be identified correctly and with the intended name, as well as more likely to attract visual attention. As the non-referent critical objects were never named overtly, either before (in contrast to picture-word interference experiments) or throughout the experiment, we believe that the necessity for subjects to correctly identify especially phonologically related critical objects justifies this decision.

[9] As there were 30 items overall, the critical object's position within the grid (which had four spaces) could not be balanced perfectly, so two grid positions contained seven items' critical objects and two positions contained eight items' critical objects

The spoken stimuli were recorded by a female native speaker of standard German. The average duration of a spoken word was 613 msec. As the exact same spoken stimulus (e.g. "Gewehr") was used for all conditions of a given item, there was no need of controlling these durations further.

Thirty experimental items and 30 filler items were constructed. On all filler items the spoken stimuli were non-words instead of words.[10] In six of the filler items the spoken non-word's onset overlapped with one of the depicted objects' names.[11] In addition, there were ten practice items that mirrored the types of items used in the experiment itself.

Pre-tests We conducted both naming and relatedness pre-tests for our stimuli. In the naming tests, participants were presented with our visual objects one at a time and were requested to give a single name for each one. In the rating tests, participants saw a visual object and a written object name and were asked to rate the semantic relatedness of these two objects on a seven-point scale. Both types of experiments were conducted as web-experiments with German native speakers as participants. None of these participants took part in Experiments 1 to 7 described in this thesis. Having conducted one naming (22 participants) and one relatedness (38 participants) pre-test on the first version of our stimuli, we adjusted some items (re-drew some objects and re-assembled items) and then conducted a second naming pre-test (15 participants) on those improved materials.

For the 30 final items, critical objects were selected only if they were named correctly in the pre-test more than 75% of the time.[12] Semantic relatedness was tested for each phonologically related, semantically related and unrelated critical object in relation to the written version of the spoken word. In all final items phonologically related and unrelated items received relatedness ratings (on a scale from 0 to 6) below 1.2, with an average of 0.3, and semantically related objects received ratings above 3.4, with an average of 4.5 .

Design Every participant was presented with one condition of each item, and thus saw seven or eight items in each condition.[13] Each participant was assigned to one out of four different

[10] We are grateful to Andrea Weber for kindly providing these non-words, which were phonologically well-formed in German.

[11] This was done to ensure that phonological overlap between the spoken and visual stimuli would not occur on experimental items only and thus would not provide a helpful cue for the lexical decision.

[12] For the referent, the semantically and unrelated critical objects, a name was considered 'correct' if participants gave the intended or a synonymous name, as this would still ensure recognition of the object as well as semantic competition. For the phonologically related critical object only the intended name was accepted, as only this name would enable phonological competition with the spoken word in the actual experiment.

[13] Again, as there were 30 items and four conditions, the number of items a participants would see on each condition could only be approximately balanced, but each item was seen equally often in each condition across subjects.

subject lists that denoted which experimental item this participant would see in which condition. Across those four lists, each experimental item was seen exactly once in each condition. In sum, the experiment had a 4 (critical-object relatedness) x 1 within-subject design.

Procedure

Visual stimuli were presented on a 19 " computer screen with a resolution of 1024 x 768 pixels, at an average distance of 70 cm from participants. Participants' eye movements were monitored using a head-mounted SMI EyeLink I system with a sampling rate of 250 Hz. Auditory stimuli were played over an external speaker connected to the stimulus PC. The software used to present our stimuli was custom-developed in our laboratory. Participants' oral responses (as well as the auditory stimuli) were recorded separately using a Mini Disk recorder.

Upon arrival, participants were familiarized with the equipment and handed a leaflet with the written instructions for the experiment. Participants were instructed to decide, as quickly as possible, whether what they would hear with each new display was a word or a non-word. They were to do this by saying 'yes' (for having heard a word) or 'no' (for having heard something else). At the same time they were to look at the objects displayed on the screen. To motivate the presence of these objects and to draw participants' attention away from us monitoring their eye movements, we used a 'cover story': participants were told that we would measure their pupil size throughout the experiment to see how challenging they found the lexical decision task. The objects in the display were said to be necessary to have participants look at the screen, which was crucial for us to be able to film their eyes. When the participants had read these instructions, the eye-tracker was set up and calibrated, Then, participants went through the ten practice trials and were given the opportunity to ask questions about their task. Having answered these, the experiment began. Every trial, including the practice trials, had the following structure: The four objects were displayed for the entire duration of a trial. After 1050 msec[14] of presentation of these objects, the audio file began to play. Each audio file began with a 'click' sound that lasted exactly 100 msec and was necessary to determine auditory lexical decision latencies. This click immediately preceded the spoken word or non-word. Thus, the spoken stimulus began after a 1150-msec long preview of the visual stimuli. Participants responded to the auditory stimulus by saying 'yes' or 'no' as their lexical decision. Every trial had a total duration of five seconds; after that the next trial began, whether the participant had given their response or not. Each participant was presented with his or her own pseudo-randomized order of their list of trials, the only restriction on the trial order being that the first trial of the experiment was always a filler

[14] We had aimed at a preview of 1000 msec, but our stimulus presentation software took 50 msec to load an audio file.

trial. After having finished all trials, which took about 20 minutes overall, the participants were debriefed.

Analysis

All of the experiments in this thesis apart from Experiment 4 similarly combine visual-world and lexical decision setups. The questions that we want to answer with our experimental data in all of these experiments are as follows:

1. Does the spoken word (or, in later experiments, a visual cue) cause shifts of attention to any of the different types of critical objects (or highlighted objects, in later experiments)? Here, the amount of visual attention directed towards the critical object is compared to the amount of visual attention directed towards distractors (which provide the 'visual attention baseline') after the spoken word's onset. This is done separately for each condition.

2. What is the impact of the type of the critical object on the likelihood of shifting attention towards the critical object in response to the spoken word? Here, the amount of visual attention directed towards the critical object relative to the visual attention baseline after word onset is compared between conditions.

3. To what degree are the different critical object types attended to during word comprehension? This might be expected to modulate the critical objects' impact on the comprehension of the spoken word. To address this question, the absolute probabilities of participants having looked at the critical object before the lexical decision are compared between conditions.

4. What is the impact of the type of the critical object on the speed of comprehension of the spoken word? This question refers to the comparison of lexical decision latencies across conditions.

We compiled a number of different object inspection probability measures in order to assess participants' deployment of visual attention in the display as well as lexical decision latencies in order to assess the speed of comprehension of the spoken word. In the following, it will be made clear how these types of experimental data differ and address the above questions.

Data: Deployment of visual attention Inspection data were computed from the raw EyeLink output, which for each inspection made on a trial identifies which object on the screen (critical object, one of the distractors, or background) was looked at as well as the start and end times of that inspection relative to trial onset.[15] An inspection comprised any number of successive fixations[16] on the same object, so that the start of the inspection marked the start of the first fixation to that object after another object had been looked at, and the end of the inspection marked the end of the last fixation on that object before a different object was looked at.

The amount of visual attention an object received was always expressed in terms of an *inspection probability*. The inspection probability for a certain object and time region on a given trial is either one if there was at least one inspection on that object during that time region or zero if there was not. When these inspection probabilities are averaged, e.g. for one participant, they indicate the proportion of trials out of all that participant's trials on which the object was looked at during that time region.

For our statistical analyses, several visual attention measures were computed with the following properties that varied depending on the specific current question:

Time regions: before shifted word onset, during the shifted spoken word, and before the lexical decision: As the onset of the spoken word was expected to affect the amount of visual attention that objects received over time, we calculated our visual attention measures for different time regions over the course of a trial.

It is important for these time regions to have similar average durations across conditions, because longer time regions provide more of an opportunity to look at any given object. If time regions' durations co-varied with the experimental manipulation, this could contaminate inspection probabilities. We chose three time regions that fulfilled this requirement: The time region *before word onset* (i.e. from trial onset to word onset),[17] the region from word onset until word end,[18] and the region spanning the one second before the lexical decision was made.[19] As both word end and the lexical decision happened at different points in time relative to word onset on every trial, these last two time regions' boundaries were determined individually on a trial-by-trial basis. This had the consequence that the spoken input actually perceived during the time region of the second before the lexical decision could differ slightly from trial to trial. Given the requirement of a time region with a constant average length across conditions, and

[15] In Experiment 1, the area of the display was divided into four squares which served as areas of interest for determining which displayed object was looked at during an inspection.

[16] Gaze data was summarized into fixations by the EyeLink software, a fixation being defined by gaze hitting the screen in about the same location for at least 100 msec. We used those fixation data without adding adjoining blinks or saccades to them.

[17] This time region had the exact same onset and offset and lasted 1150 msec on every trial.

[18] This time region's length, the spoken word's duration, was different for every item. Across all participants, however, each item occurred equally often in each condition, so that this time region's length was again similar on average when compared between conditions.

[19] Even though this time region began at different points in time relative to trial and word onset on different trials, its duration was the same on every trial: 1000 msec.

our theoretical interest in eye movement behavior *leading up to* the end of the comprehension process of the spoken word as marked by the lexical decision, this fixed-length time region before lexical decision was preferred over a fixed-length time region starting at word onset or a time region lasting from word onset until lexical decision. Also, the one-second duration ensured, for the current experiment, that the entire word was perceived within this time region on 65% of trials (the average lexical decision takes place at around 930 msec after word onset, cf. figure 2.3).

Eye-movement data for the time region before word onset was used to determine whether the displayed objects were all equally visually attractive by themselves before the impact of additional cues (for Experiment 1, this was the spoken word). Inspection probabilities for the time region of the spoken word were used to answer questions 1 and 2, i.e. to see whether and to which degree the experimental cues caused shifts of attention to the critical object. The beginning and end of this time region were shifted by 180 msec, accounting for the approximate amount of time that it takes to direct one's gaze at an object in a visual world (Altmann & Kamide, 2004). In order to analyze the amount of visual attention received by the critical object as it might affect lexical decision latencies (question 3), we looked at the time region of the second before lexical decision, using its actual, non-shifted boundaries ('raw' lexical decision time etc.).

These time regions were used for our inferential statistics; a more fine-grained temporal impression of the deployment of visual attention can be gained from time-course graphs of visual attention (see below).

Inspections started vs. inspections lasting: When trying to find out how an experimental cue (here: the spoken word) changed the focus of participants' visual attention (questions 1 and 2), only those inspections were counted that *started* during a period of time, intending to capture the *shifts* of attention caused by the experimental cue (i.e. the spoken word and, in later experiments, a change in the display). In contrast, for hypotheses concerning the *intake* of visual information as it might affect other levels of processing (question 4), it is interesting where participants looked during a certain period of time, regardless of whether these looks started before or during that time period. So, for measures of inspections *lasting* during a time region, we counted all those inspections that started before and lasted into a time region as well as those started during that time region.

Critical-object advantage measures vs. simple inspection measures: Since a person can only fixate one object at a time, the probabilities to look at all the objects in a display are strongly intertwined.[20] The amount of visual attention received by a single object depends not only on

[20] As one can only fixate one object at a time, the amount of visual attention received by one object will always be

that object's attractiveness, but also on the attractiveness of the other objects in the display. When investigating how the attractiveness of a single visual object (in our case: the critical object) is affected by an experimental cue (questions 1 and 2), it is therefore necessary to also take into account the amount of visual attention received by objects whose attractiveness can be assumed to remain unaffected by the experimental cue. This 'visual attention baseline' is provided by our distractor objects. The three distractors' inspection probabilities were averaged into a single average distractor measure and a *critical-object advantage measure* was computed by subtracting the average distractor data from the critical object data. When there were more looks to the critical object than to the average distractor, this critical-object advantage would be positive, when there were more looks to the average distractor, the critical-object advantage would be negative. In contrast, when we were interested in the probability with which the critical object was looked at as the potential source of an effect on the comprehension of the spoken word (question 3), *'simple', absolute, critical-object inspection probability measures* were used. That is, the actual proportions of trials out of all trials on which the critical object was looked at was deemed to be more relevant to the question of whether the critical object was looked at enough to have an effect on word comprehension than the degree to which the critical object received more visual attention than the distractors.

Time course data: In order to gain an impression of the way in which visual attention was deployed during the course of an average trial, we compiled time course graphs of visual attention. The probabilities of inspections on screen objects having lasted during 50-msec time slots relative to the onset of the spoken word were calculated and plotted separately for each condition. These graphs illustrate the probabilities that given objects are looked at at different points in time - shifts of attention can be seen in the time course graphs' slopes.

Data: Lexical decision latencies Oral lexical decision latencies were hand-coded using the Praat audio file analysis software: a raw lexical decision latency was the time difference from the onset of the 'click' right before word onset until the beginning of the participant's response. For analyses of the spoken word's speed of comprehension (question 4), we calculated lexical decision latencies relative to the spoken words' offset, so as to remove some of the item-specific variation in these data.[21]

affected by the amount of visual attention received by other objects.

[21] As we were not interested in comparing lexical decision latencies between words but only ever compared lexical decision latencies for the same words between conditions (which were created by manipulating the visual stimuli), it was not necessary to determine lexical decision latencies relative to the spoken words' uniqueness points (Goldinger, 1996) or otherwise take into account the spoken words' characteristics in the lexical decision analyses.

Data removal Before analysing the data, it was necessary to remove some of it. The procedure used consisted of several steps and was the same for all experiments described in this thesis, save for specific parameters that will be made explicit for every single experiment:

Removal of incorrect-response trials: First, all trials with incorrect lexical decision responses (eleven trials in Experiment 1) and all those with the response given before word onset were removed (three trials in this experiment). In Experiment 1, the resulting data set still contained 98.1% of the trials from the initial trial set.

Lexical-decision outlier removal: Following this, lexical decision data were stripped of outliers in a two-step recursive procedure: First, those lexical decision latencies whose distances from their by-participant-and-condition or by-item-and-condition means were larger than two standard deviations were removed routinely.[22] In Experiment 1, 16 lexical decision latency values were removed based on this criterion. In a second step, a number of scatterplots were created from all remaining lexical decision values: for each condition, lexical decision latencies were plotted once by participants and once by items. Based on the subjective visual inspection of these plots by the author, further few obvious outliers were identified and removed. The specific boundaries defining these outliers varied between experiments. For Experiment 1, this procedure led to two lexical decision latencies larger than 1000 msec being removed. Altogether, 2.6% of the lexical decision latencies out of those in the initial data set were excluded in Experiment 1.

Removal of subjects or items with too few critical-object inspections: Since the recognition of and a certain amount of looks to the critical objects were a precondition for our lexical decision latency analyses, we considered it necessary to eliminate the data of participants who rarely looked at the critical object as well as items for which the critical object was rarely looked at during the second before the lexical decision. The exact minimal inspection probabilities used as boundaries for this elimination were determined subjectively (upon inspection of every participant and item's average critical object inspection probabilities for the time region of the second before the lexical decision) and could differ between experiments.[23] The two participants whose data was eliminated in this step in Experiment 1 looked at the critical object during the second before lexical decision on only seven or 33% of all trials, respectively. No items' data were removed. Overall, 5.8% of the trials of the initial data set were removed in this step, resulting in a final data set that contained 89.9% of the initial, raw data set's trials and 87.3% of the initial data set's lexical decision values.

[22] In contrast to the removal of trials with incorrect responses, only the lexical decision latencies of these trials were removed, creating missing values for the lexical decision latencies of these trials but keeping the visual attention data.

[23] In this step, an important factor determining the choice of those participants or items whose data was eliminated, apart from low critical object inspection probabilities, was that their elimination did not disturb the balance of participant lists and item groups.

Subsets of trials with ideal visual attention: In order to additionally create a homogenous data set for our lexical decision analyses, a trial subset was compiled that contained only those trials on which participants had looked at the critical object during the second before they made their lexical decision, containing 65% of the cleaned-up data set's trials.

Inferential statistics To address question 1, it was necessary to compare the probabilities of inspections to the critical object and an average distractor. So as not to violate assumptions of independence in our statistical tests, we used the critical-object advantage measure described above and compared its condition means to zero in one-sample t-tests. Critical-object advantages significantly higher than zero would indicate more inspections to the critical object than to the average distractor, values significantly lower than zero indicated more inspections to the average distractor than to the critical object.

In answering questions 2 to 4 we needed to assess the impact of our critical-object relatedness factor on visual attention and reaction time variables. In the ongoing debate about the appropriate way of analyzing the data produced in psycholinguistic experiments[24], we decided, for the experiments in this thesis, to perform analyses of variance as described in Raaijmakers et al. (1999), for the reasons given below.

Experiment 1 was designed in such a way that its stimuli can be considered to be counterbalanced as defined in Raaijmakers et al. (1999): the characteristics of the different experimental materials contributed equally to the different levels of the experimental factor. The experimental items were grouped into consistent item groups (Raaijmakers and colleagues: item lists), such that for any given participant all items of that group appeared in the same condition. Participants were assigned to groups called participant lists (Raaijmakers et al: subject groups), such that the participants on the same list would see the same items in the same condition. Following a Latin-Square design, each participant saw each item once in one condition, and across participant lists every item was seen in every condition equally often. The only mis-match between our design and Raaijmakers et al's definition of a counterbalanced design was that, while we did actually use the same spoken word on all conditions of an item, part of our visual stimuli (the critical object) differed systematically across conditions. Since we took great care to balance crucial properties of the different types of critical object however,[25] we do treat our stimuli as if the

[24] Raaijmakers, Schrijnemakers, and Gremmen (1999); Raaijmakers (2003); Baayen (2004) and several papers in Forster and Masson (2008) discuss various alternatives to calculating by-subject and by-item ANOVAs.

[25] What we call an *item* corresponds to what Raaijmakers and colleagues call a *block* in an experimental design with matched items: the materials we used on the different conditions of an item (in Raaijmakers et al's terminology: the items of a block) are matched perfectly for the spoken stimuli (the same spoken word is used on all conditions) and on all the relevant dimensions for the visual stimuli (the different instances of the critical object, which is the only object in the display that varies across conditions, are matched for visual detail and saliency as well as name

same materials were used on every condition. Raaijmakers et al. (1999) propose that the correct analysis of variance for a counterbalanced design with a four-level within-subject factor is one with this factor as a within-subject factor and the participant list as a between-subject factor; the error terms of this ANOVA are adjusted depending on the impact of the item group factor (the exact procedure is described in Raaijmakers et al. (1999)). This is the analysis we used in order to establish the impact of the critical object's relatedness on this experiment's visual attention and lexical decision data.

2.1.2 Results

In this section, the results of Experiment 1 are reviewed. First, the analyses of different kinds of visual attention measures (as they are described in Section 2.1.1) provide insights into how visual attention was deployed before and while the spoken word was being comprehended as well as while the lexical decision was made. Then, the section closes with a review of how the accuracy and speed of lexical decision responses were affected by the visual stimuli. The results of all the experiments described in this thesis will be presented in sections structured similarly to Experiment 1's results sections. Please refer to the 'Analysis' Section of Experiment 1 as well as to the Glossary for the meaning of specific analysis-related terms that are used in the following results section.

Deployment of visual attention

Several aspects of participants' deployment of their visual attention within the display are crucial for understanding the way in which linguistic and visual processing interact in the setting presented here. In the following, each analysis is detailed in its own paragraph. To begin with, the development of inspection probabilities for all displayed objects over the course of a trial is presented in order to get a first impression of how visual attention was deployed within the display relative to the onset of experimental cues and to the lexical decision. This is followed by an analysis that intends to uncover whether any of the displayed object received more attention than others before the onset of the spoken word (or, in later experiments, visual cues) on any condition, which they should not. Potential fixation preferences during this time region are crucial to know about, as they would create an imbalance in the preconditions for the subsequent effects of the experimental cues on participants' viewing behavior. In the paragraph after that, looks to the critical object and the visual baseline object(s) (i.e. the distractors) are assessed for each condition as well as compared across conditions, in order to find out whether participants shifted their

properties, cf. Section 2.1.1).

visual attention towards the critical object in response to the onset of the spoken word. Lastly, the likelihood to have looked at the critical object during the second before the lexical decision is compared across conditions. This analysis should ideally show that the critical object received enough visual attention during the spoken word's comprehension on all conditions that lexical decision latencies could actually potentially be affected by the processing of the critical object.

Time course of visual attention: Overview Figure 2.3 provides a first impression of the way in which visual attention was deployed over the course of an average trial in each of the four conditions.[26]

At word onset, the probabilities to look at the critical object and at the average distractor should ideally be equal, as all objects in the display should be equally visually attractive until the spoken word's starts directing looks. For our data, this seems to hold approximately for all conditions, as critical object and average distractor inspection probabilities diverged only slightly. After word onset, a rise in the probability to look at the critical object can be seen clearly only on the referent condition, where looks to the critical object began around word end and peaked around the time of lexical decision. There was an early, weak tendency to look at a phonologically related critical object during the spoken word and a later, only slightly stronger numerical tendency to look at the semantically related critical object, which still peaked before the lexical decision is made. The unrelated critical object was not looked at more than an average distractor in response to the spoken word.

Assessing the visual saliency of the displayed objects: Critical-object inspection advantages before word onset Analyses of critical-object advantages for inspections started during the 'shifted time region before word onset' (trial onset + 180 msec until word onset + 180 msec) revealed slight imbalances between the amounts of visual attention directed at the critical objects and at the average distractor before word onset (cf. Figure 2.4). One-sample t-tests comparing these critical-object advantages to zero for every condition showed a marginally significant tendency to inspect the referent critical object more than the average distractor before word onset ($t_1(21) = 2.2$, $p < .05$, $t_2(29) = 1.1$, $p = .28$), while the semantically related critical object was looked at marginally less than the average distractor ($t_1(21) = 3.3$, $p < .01$, $t_2(29) = 1.4$, $p = .17$). There were no significant differences in inspection probabilities for the critical and

[26]Confidence intervals in these graphs are calculated as CI(95%) = $\bar{x} \pm 1.96 * SE$, where $\bar{x}$ is an average inspection probability and SE that probability's standard error. They are included to provide an impression of the approximate time regions during which the tendencies to look at different objects might diverge significantly (as is addressed in later parts of the Results Section), but not in order to make claims about the significance of such differences for specific time slots.

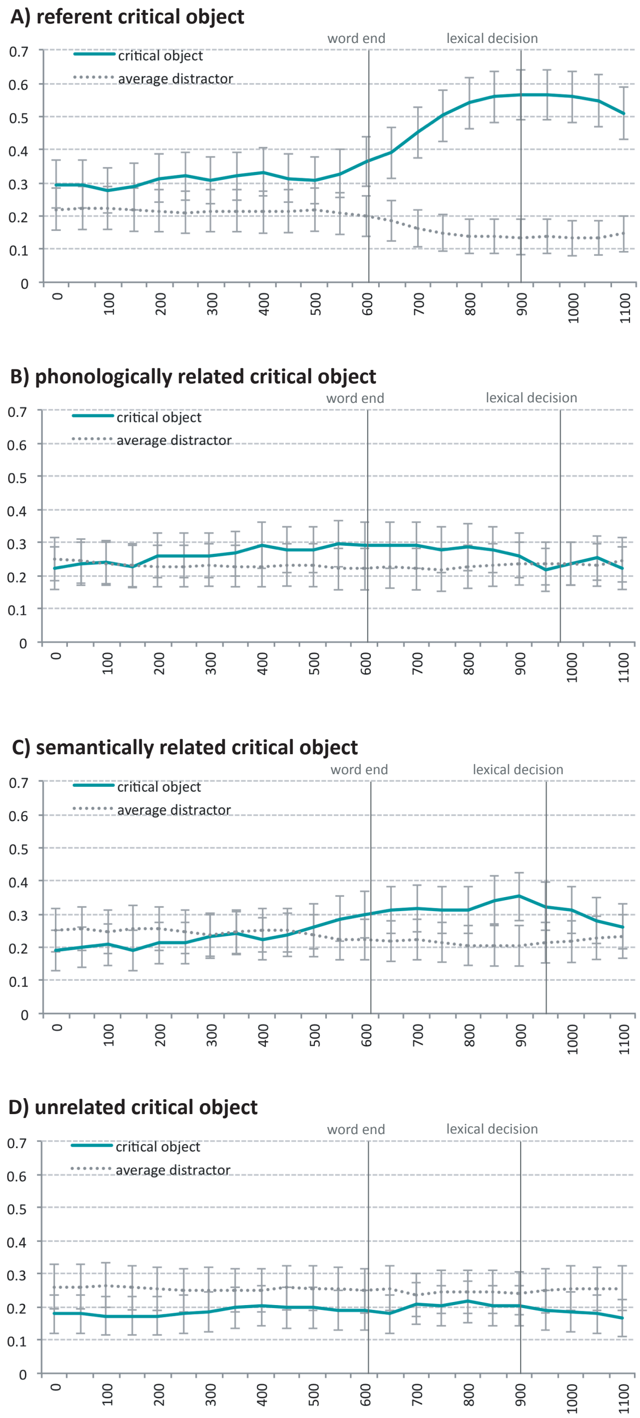

Figure 2.3: Experiment 1: Critical-object and average-distractor inspection probabilities for 50-msec long time slots relative to the onset of the spoken word, with one graph (A to D) per condition. The time identifying a time slot denotes its end point. Error bars indicate 95% confidence intervals.

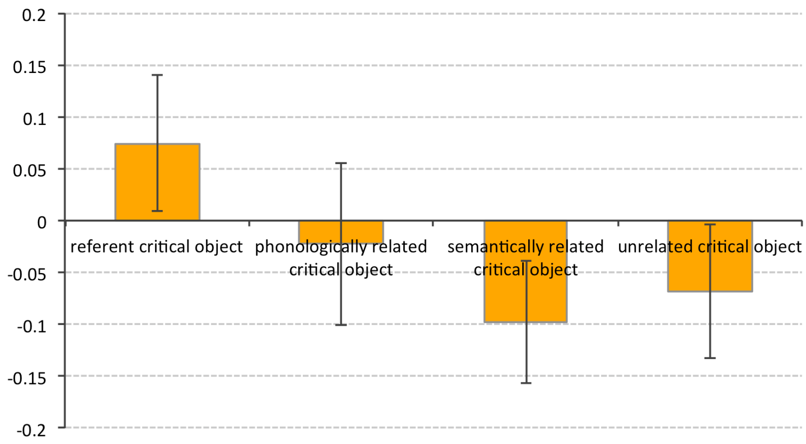

Figure 2.4: Experiment 1: Average critical object advantages for inspections having been started before the shifted word onset, one bar per condition. Error bars indicate 95% confidence intervals.

average distractor objects on the phonologically related or unrelated conditions (phonologically related critical object: $t_1(21) = .6$, $p = .58$, $t_2(29) = .3$, $p = .76$; unrelated critical object: $t_1(21) = 2$, $p = .051$, $t_2(29) = 1$, $p = .36$).

Assessing the amount of visual attention shifted towards the critical object in response to the spoken word: Critical-object advantages for inspections started during the spoken word We first analyzed whether there were significantly positive critical-object advantages for inspections started during the 'shifted time region of the spoken word' (word onset + 180 msec until word end + 180 msec) on any of the conditions at all (cf. Figure 2.5). One-sample t-tests conducted for each condition average revealed a significantly positive critical-object advantage for the referent condition ($t_1(21) = 6.1$, $p < .001$ (one-tailed), $t_2(29) = 4.9$, $p < .001$ (one-tailed)) and the semantically related condition ($t_1(21) = 1.9$, $p < .05$ (one-tailed), $t_2(29) = 1.7$, $p < .05$ (one-tailed)), on the other two conditions the differences between the amounts of looks to the critical object and distractor were not significant (phonologically related critical object: $t_1(21) = 0$, $p = .45$ (one-tailed), $t_2(29) = 0$, $p = .39$ (one-tailed); unrelated critical object: $t_1(21) = .5$, $p = .32$ (one-tailed), $t_2(29) = .7$, $p = .24$ (one-tailed)). A subsequent ANOVA revealed a significant effect of the critical object's relatedness on the amount of looks shifted towards it

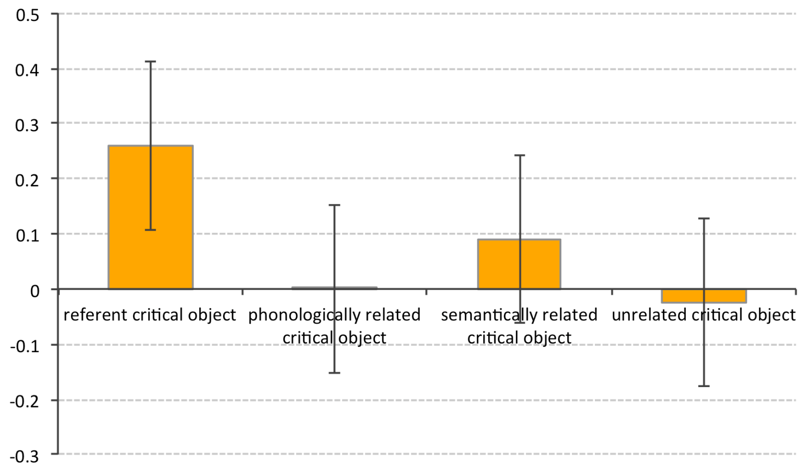

Figure 2.5: Experiment 1: Average critical-object advantages for inspections having been started during the shifted time region of the spoken word, one bar per condition. Error bars indicate 99.9% confidence intervals calculated following Masson and Loftus (2003).

during the spoken word ($F_1(3,60) = 8.8$, $p < .001$), with pairwise comparisons showing significantly higher critical-object advantages on the referent condition than on all other conditions (all $p < .01$).

Assessing the potential of critical objects to affect word comprehension: Critical-object inspection probabilities during the second prior to the lexical decision Figure 2.6 shows the probabilities that critical-object inspections had lasted during the second before the lexical decision (i.e. the per cent of trials on which there was at least one inspection on the critical object that overlapped temporally with the second before the lexical decision) for the four different types of critical objects. Participants fixated the critical object at least once during this time region on 75% of trials when the critical object was the spoken word's referent, on around 65% of trials when the critical object was phonologically or semantically related and on 55% of trials with an unrelated critical object. An ANOVA revealed a significant effect of critical object type on critical object inspection probabilities ($F_1(2,6) = 26.2$, $p < .01$), with pairwise comparisons showing significantly more looks to the referent critical object than to the other critical-object types (all $p < .01$), and more looks to the phonologically related than to the unrelated critical object ($p < .05$).

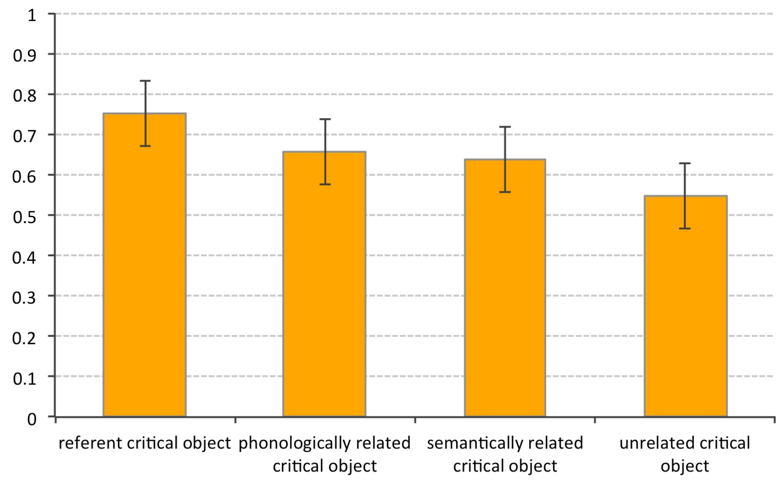

Figure 2.6: Experiment 1: Average probabilities of critical object inspections having lasted during the second before the lexical decision response, one bar per condition. Error bars indicate 99% confidence intervals calculated following Masson and Loftus (2003).

Speed and accuracy of lexical decision

The accuracy of participants' lexical decision was on average 97% (7.3 correct trials out of seven or eight per participant and condition) and did not vary significantly across conditions ($F_1(2,6) = 2.3$, $p < .25$).

Critical-object relatedness did have a significant effect on lexical decision latencies for the entire data set (cf. Figure 2.7 A); $F_1(2,6) = 29.7$, $p < .001$). In pairwise comparisons, lexical decision on the referent condition was significantly faster than on all other conditions (all $p < .05$), and lexical decision latencies on the phonologically related condition were significantly longer than on the unrelated condition (p < .05). When only those trials with at least one critical object look having lasted during the second before lexical decision (cf. Figure 2.6) were considered for lexical decision latency analysis (cf. Figure 2.7 B)), the picture remained roughly similar: there was a significant effect of critical object type, ($F_1(3,6) = 10.3$, $p < .01$) and lexical decision was faster when the critical object was a referent than when it was the phonologically related or unrelated object (both $p < .05$). Lexical decision latencies on the phonologically related condition were numerically longest, but differed significantly only from the semantically related condition ($p < .01$).

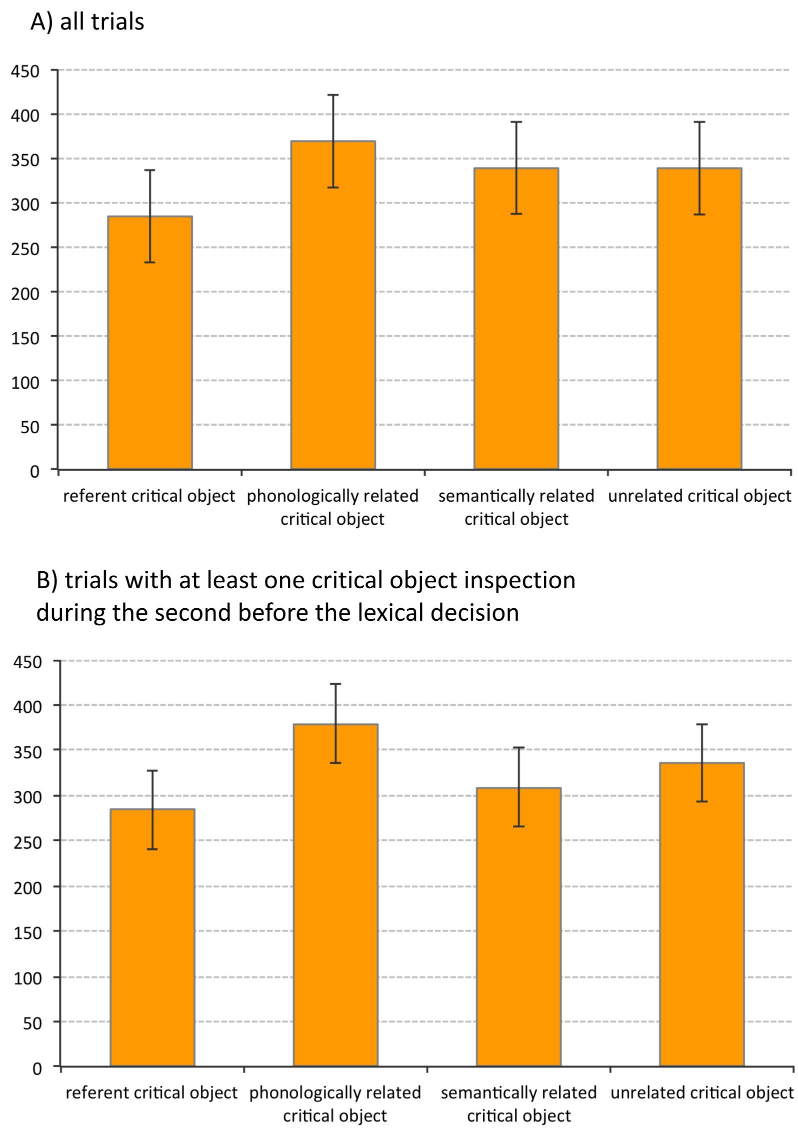

Figure 2.7: Experiment 1: Average lexical decision latencies in msec measured from word end, one bar per condition. Graphs A) and B) show averages for the entire data set and a trial subset, respectively. Error bars indicate 99.9% (Graph A) and 99% (Graph B) confidence intervals, following Masson and Loftus (2003).

2.1.3 Discussion

In this experiment, visual attention was deployed as follows: While a referent critical object was looked at reliably in response to the spoken word, the tendency to look at a semantically related critical object is significant but not very strong, and shifts of gaze towards phonologically related critical objects are visible in time course graphs but were not frequent enough to reach significance. On the basis of findings from previous visual world studies, one may have expected stronger tendencies to look at both phonologically and semantically related objects in response to a spoken word (cf. Section 1.3). That this was not found in the current experiment may be attributed to two types of viewing imbalances before the onset of the spoken word in our experiment: Firstly, the semantically related critical object was looked at less than the average distractor before the spoken word's onset, making it harder for a subsequent rise in critical-object inspection probabilities compared to average distractor inspection probabilities (i.e. a tendency to shift visual attention towards the critical object) to gain significance.[27]

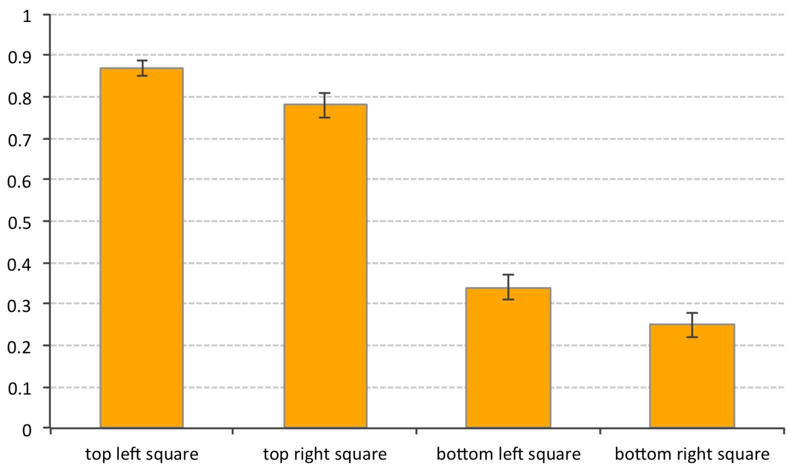

Figure 2.8: Experiment 1: Average probabilities to inspect the displayed objects before word onset, one bar per screen square (i.e. per screen position of the displayed object). Error bars show 95% confidence intervals.

Also, as is illustrated by Figure 2.8, participants preferred to scan through the four displayed objects from top-left to bottom-right (as was presumably invited by the objects' arrangement in a square: Dahan, Tanenhaus, and Salverda (2007)), which led to the objects displayed in the

[27] It could be concluded that the semantically related visual objects used in this experiment had properties that made them less attractive for visual attention than the other critical object types. However this is not the case, as no such differences in inspection probabilities during the preview period have occurred in subsequent experiments using the same stimuli. So, this effect is attributed to chance.

lower half of the display not having been looked at at all before word onset on over 60% of trials. Not having had the opportunity to visually process the critical object sufficiently before word onset on a substantial portion of trials may have prevented the recognition of the relation between the spoken word and the critical object (Huettig & McQueen, 2007), which would in turn decrease the probability to launch inspections to all three related types of critical objects. It is also conceivable that, as the relationship between the referent and the spoken word is more obvious than the relationship between the phonologically related or semantically related critical objects and the spoken word, shifts of attention towards the referent would be affected less by such a pre-word viewing imbalance than shifts of attention towards the phonologically or semantically related critical objects, matching our results. Finally, Experiment 1's task explicitly disconnected the inspection of the display from the comprehension of the spoken word: the lexical decision task was introduced to participants as the main focus of the experiment, while the display and eye tracking were presented merely as instruments to assess the difficulty of the lexical decision by measuring participants' pupil size. Previous visual-world experiments that showed NCEs to phonologically or semantically related objects had tasks that either explicitly linked display and spoken stimulus (e.g. in a task to pick up or touch the named objects (Allopenna et al., 1998; Yee & Sedivy, 2006)) or that left the linkage between images and utterances implicit (e.g. the instruction for participants to 'listen to the sentences carefully' and 'look at whatever they wanted' in Huettig and Altmann (2005)). If participants of Experiment 1 focused on the lexical decision task and tried to not be confused by the displayed objects, this may have caused them to not make the connection between the spoken word and semantically or phonologically related critical objects.

Yet another account for the combination of the observed phonological interference and the lack of NCEs towards the phonologically related critical object would be that participants avoided looking at the phonologically related critical object *because* it would interfere with the spoken word's comprehension. There are two problems with this idea though: One is a methodological concern, saying that our experimental data give us no indication why participants may have behaved in a certain way or whether there was any causal relation between participants' viewing behavior and their lexical decision latencies at all. The other one would be that the phonologically related critical object *did* interfere with the lexical decision task, so either participants did not look at it for an entirely different reason than to avoid that interference, or their strategy of avoiding interference by forgoing NCEs did not work.

To sum up our results of this first pilot study, we observed that participants chose to look at the referent of a spoken word upon recognizing it, i.e. before giving their lexical decision response. They were also faster in making the lexical decision when the display contained a referent than when it did not. While a semantically related critical object attracted some visual attention but did not affect lexical decision latencies, a phonologically related critical object did not attract visual attention but slowed down lexical decision.

2.2 Second pilot study: Experiment 2

The goal of Experiment 2 is similar to Experiment 1's goal: to explore the interaction of word comprehension and the deployment of visual attention when the visual environment contains objects that can affect the word comprehension process in beneficial or detrimental ways. As discussed above, several aspects of Experiment 1's method presumably discouraged NCEs: the instruction obscured the connection between displayed objects and spoken noun and the square display layout put the objects in the bottom at a disadvantage for being recognized. For Experiment 2, the method of Experiment 1 was modified slightly (cf. Figure 2.9) in order to ease the recognition of especially the phonologically and semantically related critical object types, thus encouraging voluntary noun-contingent eye movements towards them upon comprehension of the spoken word as well as increasing their impact on the word comprehension process. This was implemented by displaying only three instead of four objects, by using a longer preview and by arranging the displayed objects in a way that would discourage a reading-like scanpath during the preview period. Also, the instruction given to participants did not explicitly disconnect visual and linguistic stimuli anymore and lexical decision responses were now not given orally, but via button presses.

Again, as Experiment 1, Experiment 2 is a pilot study for the empirical questions tackled in this thesis. This means that there are no specific hypotheses for this experiment but rather the questions of how likely participants are to fixate different types of critical objects related to a spoken noun as well as how these objects affect the noun's comprehension.

2.2.1 Method

Participants

Thirty-two German native speakers from the Saarland University community with normal or corrected-to-normal vision took part in this experiment and were paid five Euros each. The data of eight participants was excluded from the analyses because their gaze could not be calibrated well, they turned out to be bilingual, they did not comprehend the task, or (in one case) they appeared nervous and distracted throughout the experiment. Overall, 24 participants' data was analyzed.

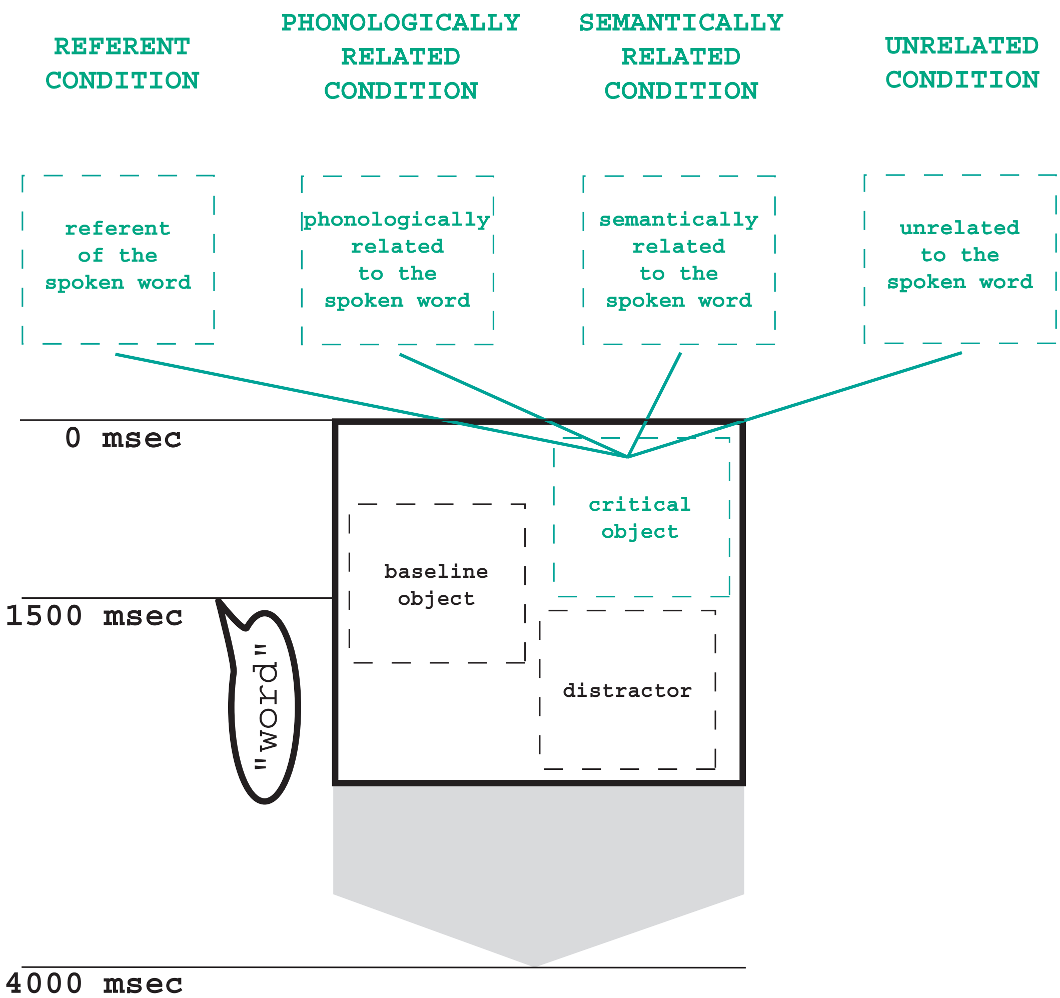

Figure 2.9: A sample trial of Experiment 2. The display, containing the critical object (the spoken word's referent / a phonologically related object / a semantically related object / an unrelated object) and two distractors, was shown with a preview of 1500 msec. Then, the spoken word was played while the display remained unchanged. Participants gave their lexical decision response via a button press. Every trial had an overall duration of four seconds.

Materials

This experiment used 28 items, which were a subset (see below) of the same materials as Experiment 1 (for a sample stimulus, see Figure 2.10). To counteract the effect of participants' tendency to scan the display along a top-left to bottom-right reading-like scanpath, the display layout was changed for this experiment. Now each display consisted of the critical object and only two distractor objects, arranged in a triangular layout. The single objects each fitted into a 130 x 130 px square, which stretched across 4 °of visual angle.

Every experimental item's display was created using one of seven possible triangular layouts, which differed only in the rotation of the (virtual) equilateral triangle on whose corners the objects were located. For every layout there was one designated critical object position, and across all seven layouts this position was distributed evenly across all regions of the display. None of the layouts had a rotation which aligned any of its edges with a screen edge. These measures ensured both that the three objects on the screen were never arranged in a 'conventional' grid-like way and that the layout of the display was constantly changing, discouraging a possible preference in participants to scan the screen in a reading-like fashion.[28]

The spoken words, recorded by a female German native speaker, were taken from Experiment 1 and had an average duration of 619 msec across experimental trials. In order to match the number of items in each condition, only a subset of 28 of Experiment 1's experimental items were used, each of them reduced by one of their distractor objects. These objects were used to create additional filler items, so that the overall item set contained 56 filler items consisting of 14 word and 42 non-word trials for six of which the spoken non-word's onset overlapped with one displayed object. There were ten practice items mirroring all item types that occurred in the actual experiment.

Design This experiment had a 4 (critical object relatedness) x 1 within-subject Latin Square design, with four participant lists and four item groups. For this and all following experiments, participant lists and item groups were fully balanced (each experimental condition was seen equally often by each participant and for each item) unless otherwise noticed.

Procedure

The equipment used in this experiment was the same as in Experiment 1 with the exception of the stimulus presentation and data collection software used, SR Research's Experiment Builder, and a Cedrus RB-834 response pad for the collection of button-press lexical decision responses.

[28]Or, in any case, minimizing the effect such a reading-like scanpath would have on our visual attention data.

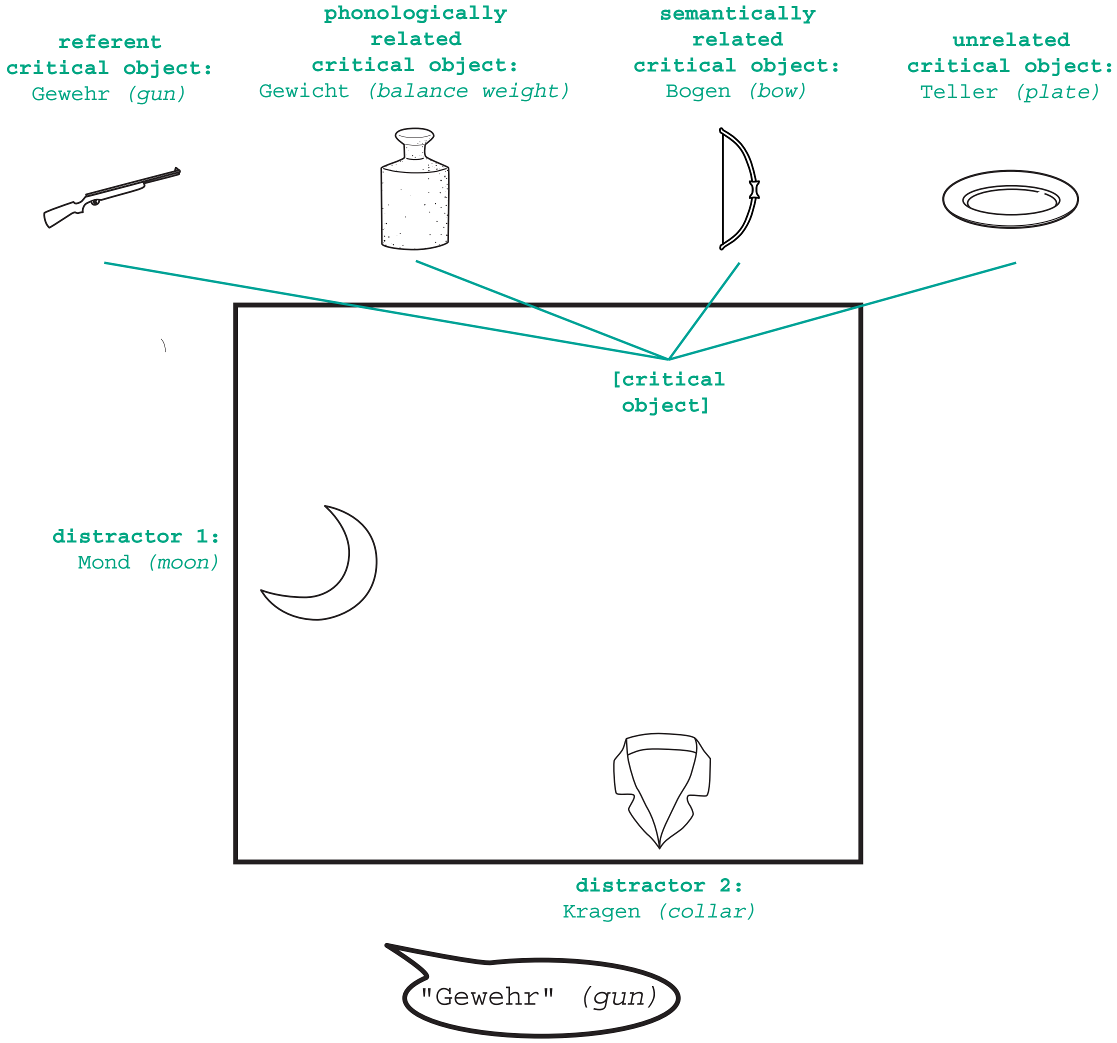

Figure 2.10: The materials for one of Experiment 2's items. Participants were presented with a display containing three inanimate objects: the critical object and two distractors. The critical object's identity was determined by the condition: it could either be the spoken noun's referent, a phonologically related object, a semantically related object or an unrelated object.

Participants received an instruction similar to Experiment 1's, apart from an explicit additional request to inspect all three displayed objects prior to the spoken word's onset and a new response modality: The lexical decision response was to be given by pressing one of two large buttons in front of the participant (both on the Cedrus response pad), with the left button to be pressed using the left hand and the right button using the right hand. Instructions were adjusted on-line so that the 'word' response was always given using the participant's dominant hand.

On every trial the display was shown for 1500 msec before the spoken word's onset and remained unchanged for the entire trial duration of four seconds.

As this experiment had twice as many filler trials as experimental trials, every experimental trial was preceded by two filler trials. The exact trial order was pseudo-random (random assignment of items to the pattern of two fillers followed by one experimental item) and new for every subject.

Analysis

The strategy used to analyze this experiment's data, including questions, measures and statistical tests, was exactly the same as for Experiment 1.

Data removal Incorrect trials and outliers were removed in the same fashion as in Experiment 1: Six per cent of the trials from the initial data set were removed because of incorrect or too early responses. Lexical decision latencies were removed for six trials with objectively determined lexical decision outliers and for three further trials trials whose lexical decision latencies were still subjectively considered outliers after the inspection of scatter plots (they had values larger than 900 msec), which led to the removal of 1.3% of data points from the initial set. As participants inspected the critical objects rather reliably, no further data was eliminated based on participants' reluctance to make noun-related eye movements. The final data set thus contained 94% of the initial data set's trials, and 92.7% of the initial data set's lexical decision latency values. The trial subset on which participants did look at the critical object at some point during the second before lexical decision, used only in the analysis of lexical decision latencies, contained 75.4% of the trials from the original data set.

2.2.2 Results

The results section of this experiment is structured similarly to Experiment 1's results section. Please also refer to the 'Analysis' Section of Experiment 1 as well as to the Glossary for the meaning of specific analysis-related terms that are used in the following results section.

Deployment of visual attention

Time course of visual attention: overview Figure 2.11 illustrates how visual attention was deployed after the spoken word's onset. In contrast to Experiment 1, there were now no imbalances between looks to the critical object and average distractor at word onset for any condition. The time curves suggest that there were shifts of attention towards referent, phonologically related and semantically related critical objects in response to the spoken word, but none to unrelated critical objects. This tendency to look at critical objects began about 100 msec before the average word end on the referent condition, at word onset on the phonologically related condition, and at the average word end on the semantically related condition. Probabilities of fixating critical objects peaked before or around the average lexical decision on all conditions, reaching the highest peak on the referent condition.

Assessing the visual saliency of the displayed objects: Critical-object inspection advantages before word onset Before word onset, one-sample t-tests comparing critical-object advantages to zero showed no significant differences between critical-object and average-distractor inspection probabilities on any condition (referent critical object: $t_1(23) = .9$, $p < .36$, $t_2(27) = .8$, $p = .45$; phonologically related critical object: $t_1(23) = 1.1$, $p < .3$, $t_2(27) = 1.1$, $p = .3$; semantically related critical object: $t_1(23) = .9$, $p < .38$, $t_2(27) = .5$, $p = .59$; unrelated critical object: $t_1(23) = 1.2$, $p < .24$, $t_2(27) = 1.1$, $p = .29$).

Assessing the amount of visual attention shifted towards the critical object in response to the spoken word: Critical-object advantages for inspections started during the spoken word Shifts of visual attention towards the critical object during the spoken word are plotted in Figure 2.12. One-tailed one-sample t-tests showed significantly positive critical-object advantages for inspections started during the 'shifted time region from word onset until word end' in the referent ($t_1(23) = 4.2$, $p < .001$ (one-tailed), $t_2(27) = 5.2$, $p < .001$ (one-tailed)) and phonologically related ($t_1(23) = 1.8$, $p < .05$ (one-tailed), $t_2(27) = 1.8$, $p < .05$ (one-tailed)) conditions, as well as a marginally significant critical object advantage in the semantically related condition ($t_1(23) = 2.1$, $p < .05$ (one-tailed), $t_2(27) = 1.3$, $p = .11$ (one-tailed)). The unrelated-critical-object advantage did not differ significantly from zero ($t_1(23) = .9$, $p < .2$ (one-tailed), $t_2(27) = .7$, $p = .25$ (one-tailed)). An ANOVA revealed a significant effect of the critical-object relatedness ($F(3,6) = 10.4$, $p < .01$). Pairwise comparisons between the single conditions showed significantly higher critical-object inspection advantages in the referent than in the semantically related and unrelated conditions, as well as higher critical-object inspection advantages in the phonologically related than in the unrelated condition (all $p < .05$).

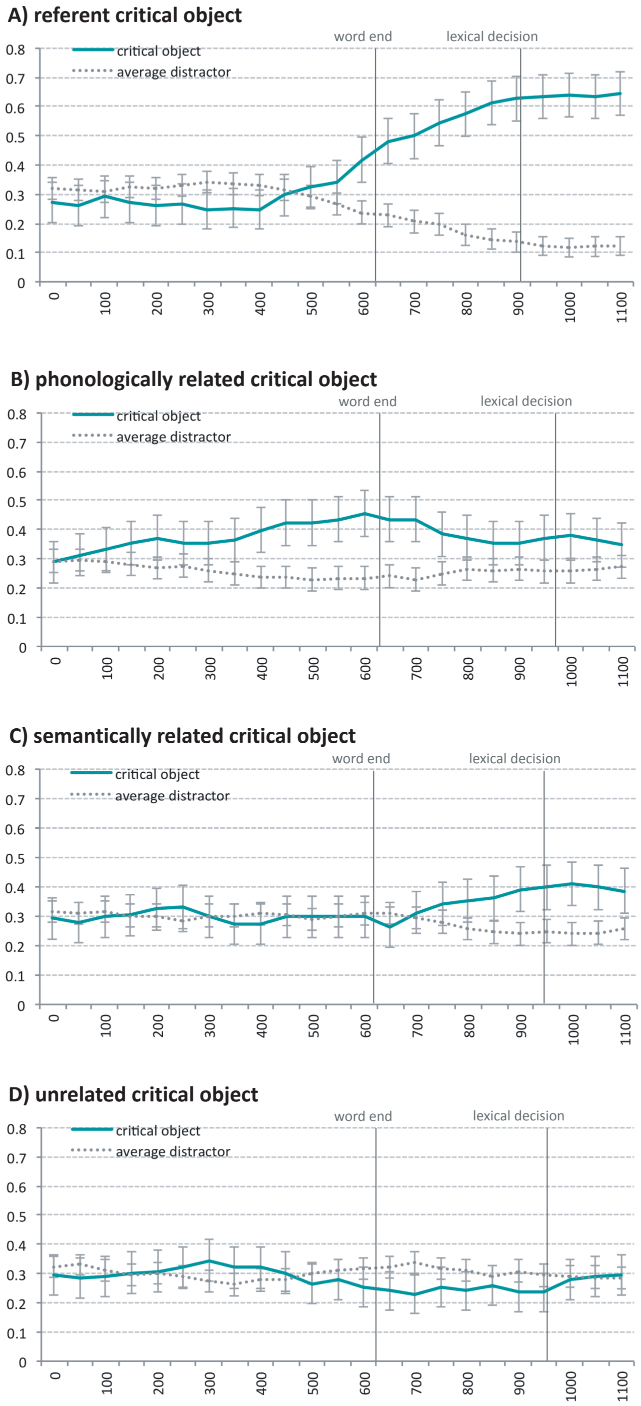

Figure 2.11: Experiment 2: Critical-object and average-distractor inspection probabilities for 50-msec long time slots relative to the onset of the spoken word, with one graph (A to D) per condition. The time identifying a time slot denotes its end point. Error bars indicate 95% confidence intervals.

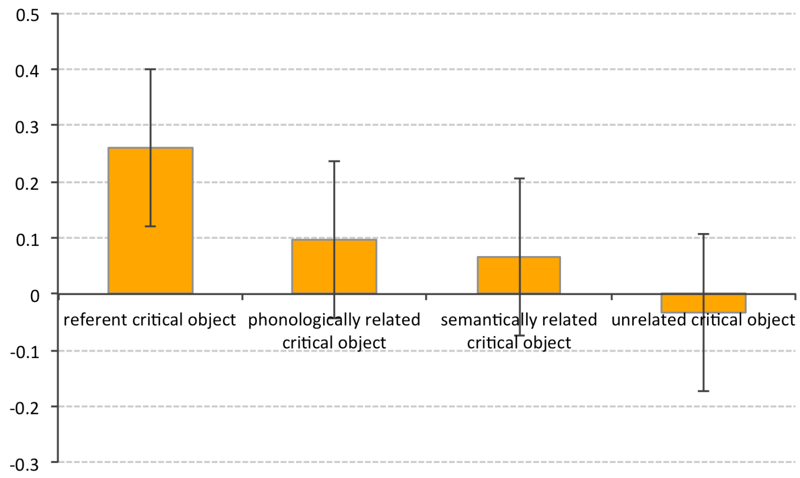

Figure 2.12: Experiment 2: Average critical-object advantages in probabilities of inspections having been started during the shifted word region, one bar per condition. Error bars indicate 99% confidence intervals calculated following Masson and Loftus (2003).

Assessing the potential of critical objects to affect word comprehension: Critical-object inspection probabilities prior to lexical decision As can be seen in Figure 2.13, the relatedness of the critical object significantly affected the probability that the critical object was looked at at some point during the second before lexical decision ($F(3,6) = 6.5$, $p < .05$): There were significantly more looks to each of the three related critical object types than to the unrelated critical object (all $p < .05$), but no differences between referent, phonologically related and semantically related critical object inspection probabilities.

Speed and accuracy of lexical decision

An ANOVA on the average number of accurate lexical decision responses (6.6 out of seven per participant and condition, which equals 94%) showed no influence of the critical object's relatedness on participants' performance in the lexical decision task ($F(3,6) = .7$, $p = .1$).

Lexical decision latencies (Figure 2.14) were affected significantly by the relatedness of the critical object for both the entire data set ($F(3,6) = 5.9$, $p < .05$) and for the data subset that contains only those trials with at least one critical-object inspection having lasted during the second before lexical decision ($F(3,57) = 4.8$, $p < .01$). In both data sets, pairwise comparisons

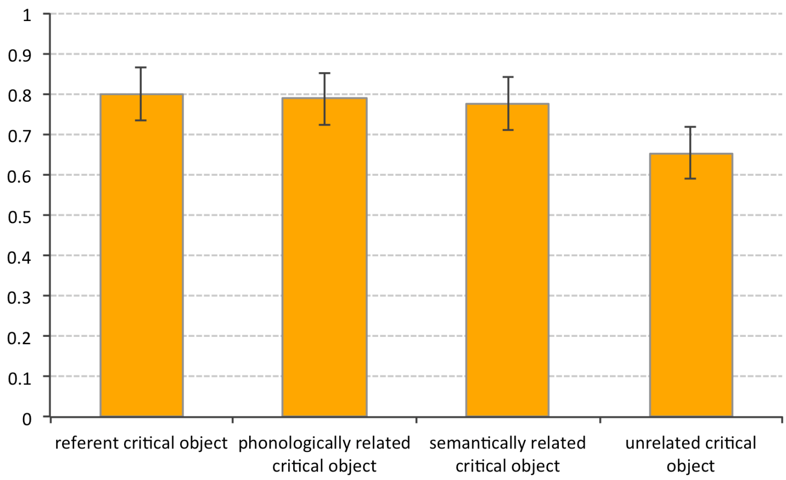

Figure 2.13: Experiment 2: Average probabilities of critical-object inspections having lasted during the second before the lexical decision response, one bar per condition. Error bars indicate 95% confidence intervals calculated following Masson and Loftus (2003).

showed significant differences to the lexical decision latencies in the unrelated condition only for the referent condition (both $p < .01$).

2.2.3 Discussion

As can be seen from Experiment 2's visual attention data, the switch from a square to a 'rotating' triangular object grid in the display eliminated the differences between critical-object and distractor inspection probabilities before the onset of the spoken word, when the critical object and the distractors should still be equally attractive to look at. More importantly, this enabled the recognition of all three related critical object types, as proved by significant tendencies to shift gaze towards the referent, the phonologically related object and the semantically related object in response to the spoken word. The critical object was looked at at least once during the second before lexical decision on most of the trials (around 78%) of the three conditions with related critical objects, and on fewer trials (still above 65%) of the unrelated condition, creating sufficiently balanced preconditions for the critical object to affect the word comprehension and lexical decision process.

Despite all types of related critical objects having been looked at during word comprehension,

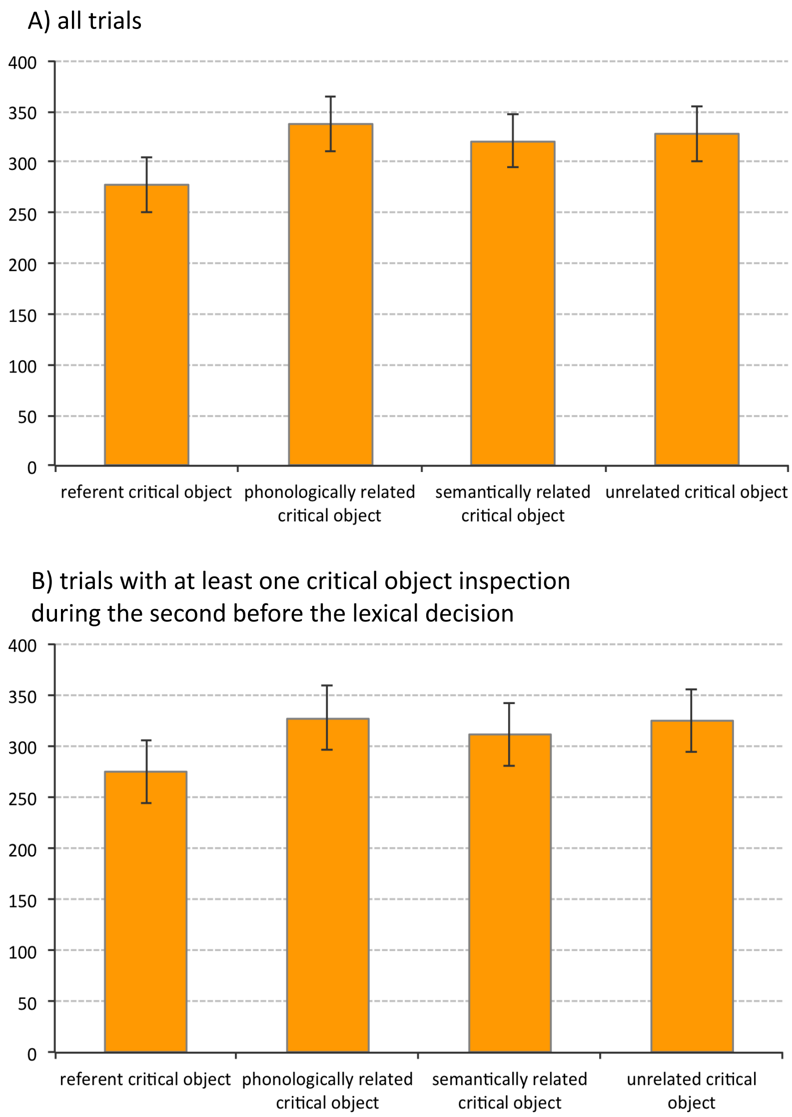

Figure 2.14: Experiment 2: Average lexical decision latencies in msec measured from word end, one bar per condition. Graphs A) and B) show averages for the entire data set and a trial subset, respectively. Error bars indicate 95% (Graph A) and 99% (Graph B) confidence intervals, following Masson and Loftus (2003).

the only related critical object that affected the speed of lexical decision was the referent, which made word comprehension faster than on all other conditions.

We found seemingly conflicting evidence with regard to the phonologically related critical object: In Experiment 1, this object was not looked at reliably after word onset and still interfered with word recognition, in Experiment 2 it received visual attention but did not affect word comprehension. This matches previous studies in indicating that phonological interference is a fickle effect: for example, Radeau, Morais, and Segui (1995), in contrast to Goldinger et al. (1992), do not find interference nor facilitation from auditory onset competitor primes in an auditory lexical decision task, with almost identical (20 msec vs. 50 msec) inter-stimulus intervals. Thus at this point, we cannot yet make a clear statement about the interaction of phonologically related visual objects and word comprehension. This issue will be taken up again in a more controlled manner in Experiment 3. The lack of evidence for facilitation by semantically related critical objects will be addressed in the two following experiments as well. In sum, our most reliable result across Experiments 1 and 2 is that participants look at and benefit from the referent critical object during spoken noun comprehension.

2.3 Manipulating visual attention: Experiment 3

The two previous experiments were designed as initial explorations of the question of how both visual attention and word comprehension are affected by the display of objects with a certain relation to the spoken word. An issue that we want to ultimately resolve though is whether there is an effect of looks to related objects during the comprehension of a spoken noun on that noun's speed of comprehension. Voluntary noun-contingent eye movements are triggered by word comprehension. This makes it difficult to draw conclusions about the effect of these eye movements on that same process of word comprehension: While the observation that fast lexical decision responses co-occur with high NCE probabilities could mean that NCEs improve noun comprehension, it could also indicate merely that successful noun comprehension is likely to trigger NCEs. This means that it is impossible to single out a potential effect of NCEs on noun comprehension when noun comprehension is free to affect NCEs (i.e. noun comprehension causes NCEs). Consequently, for Experiment 3, we chose to not let participants view the display as they wished, but to direct their visual attention during word comprehension by highlighting certain displayed objects, thus visually triggering eye movements similar to NCEs in target and timing. In previous studies Grant and Spivey (2003) improved performance in a problem-solving task by highlighting the relevant element of a related diagram, Richardson and Dale (2005) improved the comprehension of short stories by directing eye movements in a display during

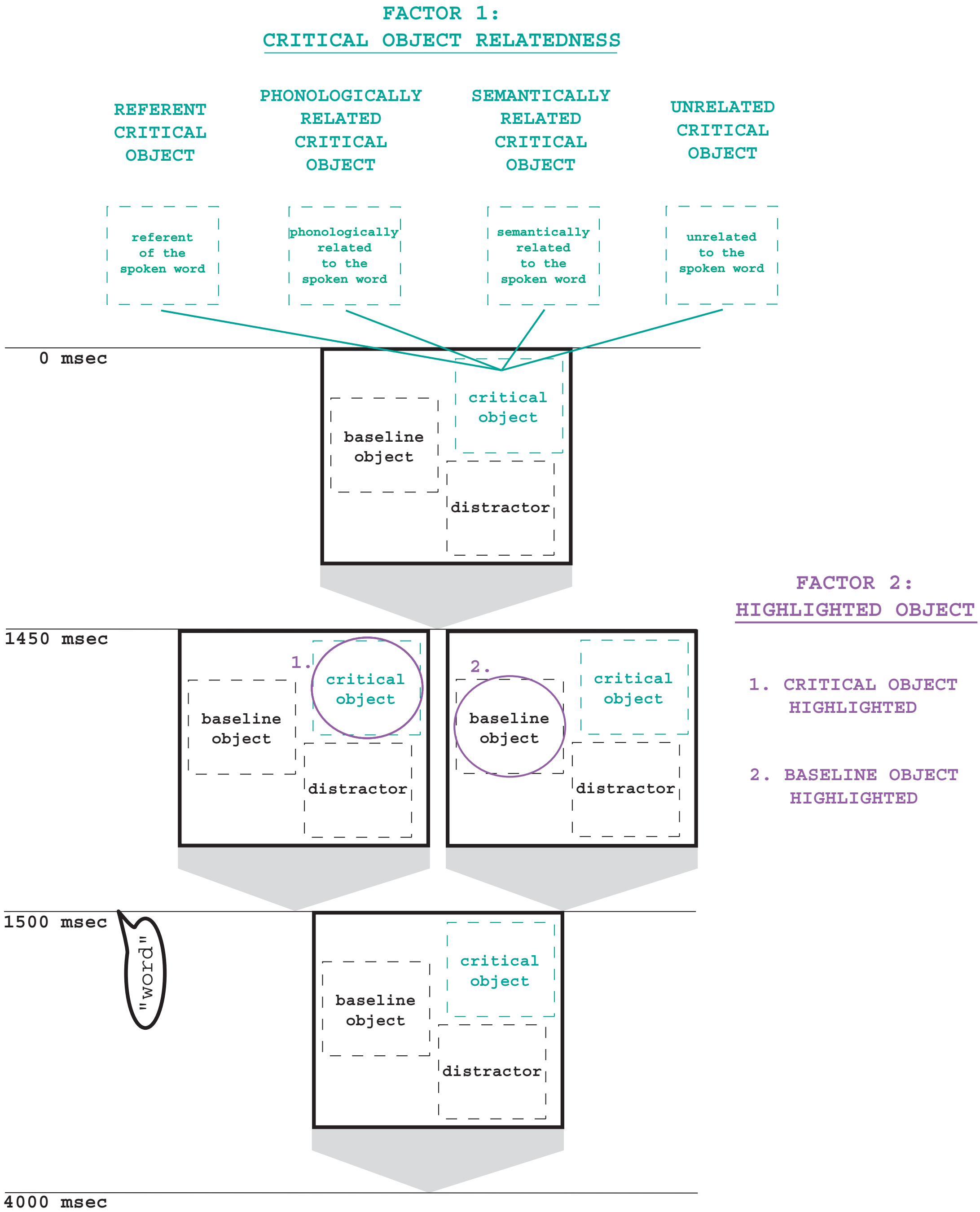

Figure 2.15: A sample trial of Experiment 3. The display, containing the critical object (the spoken word's referent / a phonologically related object / a semantically related object / an unrelated object), an unrelated baseline object and one distractor, was shown with a preview of 1500 msec. During the last 50 msec of this preview, either the critical object or the baseline object was highlighted by a surrounding circle. Then, the spoken word was played while the display still contained the three initial objects. Participants gave their lexical decision response via a button press. Every trial had an overall duration of four seconds.

fact learning in a way that they mirror the story teller's eye movements. Addressing this issue, we use low-level visual cues to direct visual attention in the display, making an eye movement to the critical object determined by the condition an item is seen in, i.e. independent of the comprehension of the spoken word. This permits conclusions about how these eye movements affect noun comprehension. In Experiment 3, we used a visual highlighting cue to attract the eye to one of the three displayed objects after the preview period (sudden onset captures visual attention: Theeuwes, Kramer, Hahn, and Irwin (1998)). This highlighted object could be the critical object or an unrelated baseline object (cf. figure 2.15).

The display in this experiment always contained three objects: the critical object (which could be the spoken word's referent, a phonologically related object, a semantically related object or an unrelated object), an unrelated baseline object and a distractor. As in the previous two experiments, the critical object was of central interest, i.e. the effect of looks to this object on noun comprehension was to be investigated. In order to separate the effect of looking at this critical object from the effect of its mere presence in the display, the unrelated *baseline object* was included in every display and visual attention was directed to this object equally often as to the critical object. In this and the following experiments, the unrelated baseline object is present in the display on every condition, is unrelated from the spoken word and may be highlighted by the visual cue just like the critical object. This creates a baseline for the analysis of lexical decision latencies: for each level of the critical object relatedness factor, the condition in which the critical object is highlighted can be compared against another condition in which the display contains the same objects, but the unrelated baseline object is highlighted. The third object in the display, the distractor, remains a baseline object for the analysis of the deployment of visual attention, similar to the previous experiments.

As Figure 2.15 illustrates, two aspects of the display were manipulated in Experiment 3: the type of the displayed critical object (the spoken word's referent, a phonologically related object, a semantically related object or an unrelated object) and whether the critical object or the baseline object were highlighted by a visual cue at word onset. Participants received the instruction to pay attention to this highlighting cue, thus we expected them to shift their visual attention towards the highlighted object at word onset reliably. As for the speed of the lexical decision to the spoken word, we expect, based on results from previous priming studies (cf. section 1.3), facilitatory effects from a referent and semantically related critical object compared to an unrelated critical object and (possibly) interference from a phonologically related critical object compared to that unrelated critical object. The main intention of this experiment is to determine how these effects are affected by whether the critical object or the unrelated baseline object has to be looked at during the comprehension of the spoken word. If these simulated noun-contingent eye movements make priming effects stronger than looks to the baseline object, this could be considered evidence for 'real-world' noun-contingent eye movements not only being caused by, but also having an effect on noun comprehension.

2.3.1 Method

Participants

Fifty-five German native speakers from the Saarland University community took part in this experiment. They had normal or corrected-to-normal vision and were paid five Euros for their participation. Data from seven participants was excluded from the analyses because of technical problems during their session, because they did not notice the highlighting cue most of the time, because they were bilingual or because they had taken part in on of the previous experiments. Data from 48 participants was analyzed.

Materials

Experiment 3 used a subset of 24 experimental items from Experiment 2, with the display containing a critical object (which could be, as in the previous experiments, the spoken word's referent, a phonologically related object, a semantically related object or an unrelated object), an unrelated baseline object and one distractor (cf. Figure 2.16). Either the critical object or the baseline object was highlighted to attract visual attention at word onset. The critical object of an item in Experiment 3 was the same as the corresponding item's critical object in Experiment 2. The unrelated baseline object, which is a concept that first occurs in Experiment 3, was selected randomly from the two distractors from Experiment 2. Its name had the same number of syllables on average as the spoken word and was slightly more frequent (by 2 frequency classes as defined by "Deutscher Wortschatz" (2007)) and it was identified reliably in naming pretests (described in Experiment 1's section above). It was selected to be unrelated to the spoken noun in Experiment 1's item selection process (as Experiment 2 - 4's stimuli were all subsets of Experiment 1's stimuli).[29]

The most important change introduced for this experiment was the highlighting manipulation (cf. Figure 2.15). While the display layout remained the same as in Experiment 2 (the only

[29]Experiment 1's semantic relatedness pretests did not test distractors, so their unrelatedness to the spoken word is verified only by the author's intuition. Consequently, Experiment 3's unrelated baseline object is unrelated to the spoken word according to the author's judgement only. Experiment 4 will show, however, that lexical decision latencies to words primed by the unrelated baseline object are the same as lexical decision latencies primed by the unrelated critical object, whose unrelatedness *was* verified in two pre-tests (cf. Experiment 1). We can thus be confident that our unrelated baseline objects are actually semantically unrelated to the spoken noun.

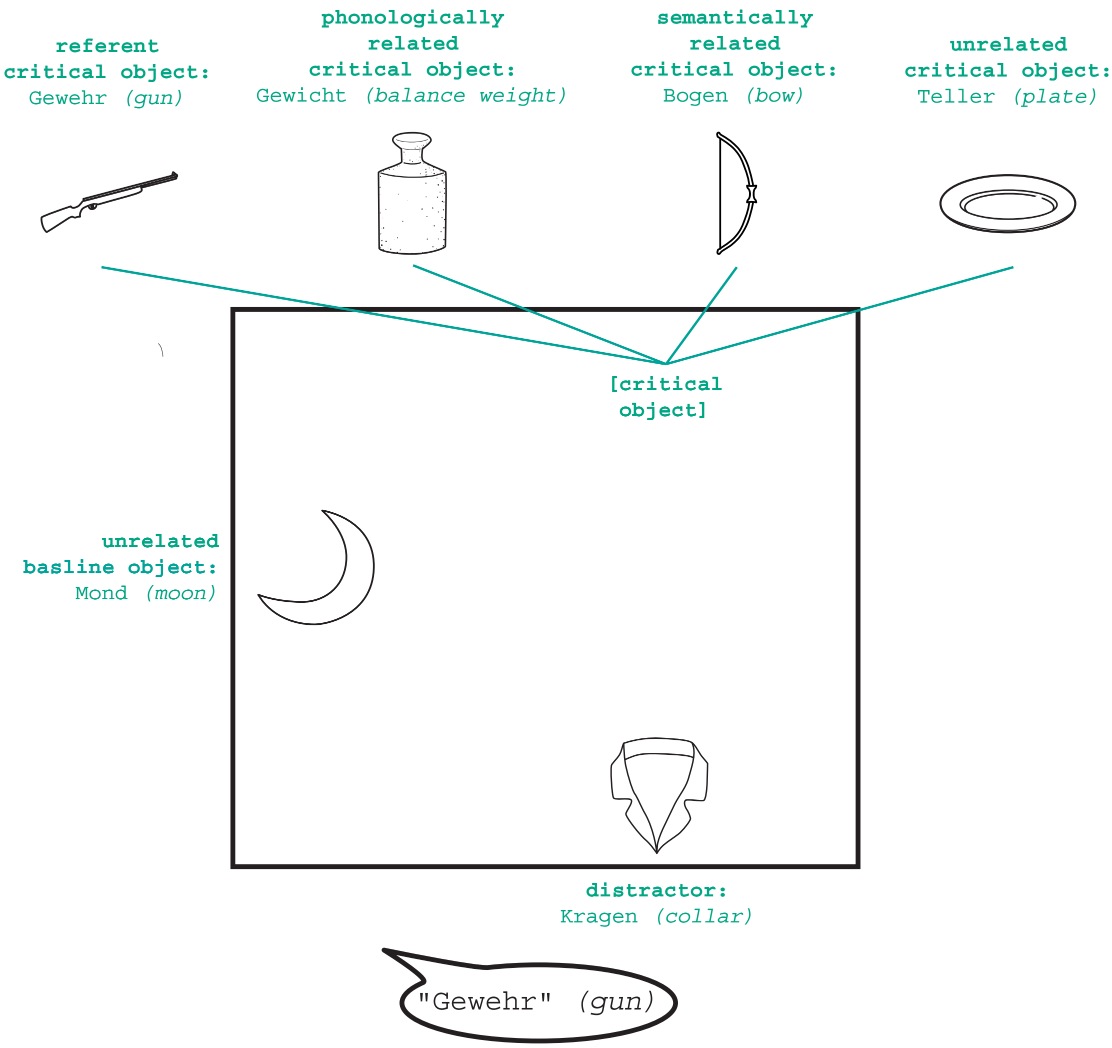

Figure 2.16: The materials for one of Experiment 3's items. Participants were presented with a display containing three inanimate objects: the critical object, an unrelated baseline object and a distractor. The critical object's identity was determined by the critical-object relatedness factor: it could either be the spoken noun's referent, a phonologically related object, a semantically related object or an unrelated object. The display remained as shown here from trial onset to trial offset, with an additional circle highlighting one object for 50 msec before word onset (cf. Figure 2.15).

change being that there were now eight instead of seven possible rotations of the triangular grid the objects were arranged in), there was now a circle being displayed around one of the objects for 50 msec right before word onset. This circle consisted of a purple outline and had a diameter of 200 pixels. Critical object and distractors of each item were positioned on the screen in a balanced fashion to ensure that potentially highlighted objects, which were critical object and baseline objects, were seen equally often in each screen quadrant across items.

The spoken words, recorded by a female German native speaker, were taken from Experiment 2 and had an average duration of 619 msec across all experimental trials.

Apart from the 24 experimental items, this experiment's trial set contained 48 filler items, out of which 36 were non-word trials (on six of them one of the displayed object's name had the same onset of the spoken non-word) and twelve were word trials without a referent in the display. On each filler trial one object was highlighted. Ten practice trials resembled the overall item set.

Design This experiment had a 4 (critical-object relatedness) x 2 (highlighted object: critical object / baseline object) within-subject Latin Square design with eight participant lists.

Procedure

The same equipment as in Experiment 2 was used in this experiment. Participants received the instruction to make their lexical decision on the spoken stimulus as fast and as accurately as possible. They were asked to look at all three displayed objects before they would hear the spoken word or non-word and to pay special attention to the object surrounded by the circle.

Each trial began with a preview of the three displayed objects, that lasted 1500 msec. During the last 50 msec of this period, the circle was displayed around one object. The circle was then removed and the spoken stimulus was played, while the three objects remained on the screen unchanged. A trial lasted 4000 msec overall.

The order of trials was pseudo-randomized for every participant, with the constraint that every experimental trial was preceded by exactly two filler trials, the second of which was always a non-word trial.

Analysis

Data The data used in this experiment, while compiled similarly to the first two experiments' data, had some new properties: Firstly, there was only one distractor object in the display that could be used as a visual attention baseline, so the distractor data displayed in the time course graphs and used in the analyses were not averages, but inspection probabilities for this one distractor only. Secondly, there were now two types of objects looks which were of interest in the analyses: the highlighted object and the critical object. The *highlighted object* is a category that does not map to one of the displayed objects, but which could be the critical object or the baseline object, depending on the condition (cf. Figure 2.15). Analyzing the amount of looks directed towards this object was necessary in order to see how well the highlighting manipulation was followed by participants. The object of interest is the critical object. It is possible that this object attracts looks independently of whether it is highlighted or not, as it may be related to the spoken word. Looks to both the highlighted object and the critical object were analyzed just like looks to the critical object in experiments 1 and 2.

Data removal After the removal of trials with incorrect lexical decision responses, there were still 97.2% of the trials left from the initial data set. As there were only three items per participant per condition in the initial data set to begin with and at least three values are necessary to compute meaningful averages, all those trials for which there existed less than two other trials from the same participant (or item) and condition were removed. The resulting data set contained 92.2% of the trials in the original data set. There were four objectively and three subjectively (larger than 1400 msec) determined lexical decision outliers, which were considered as missing values, such that 91.6% of the raw lexical decision latencies could still be used in the analyses. As the highlighted object was looked at reliably, no further data was removed.

Data subsets for the lexical decision latency analysis were not compiled because the above-mentioned few trials per participant or item and condition in the initial data set did not afford a further reduction of the data set. Also, as will become apparent in the results section, the visual cue directed participants' visual attention reliably, such that the overall data set was well-suited to lexical decision latency analyses already.

Inferential statistics It was not possible to compute ANOVAs according to Raaijmakers' instructions (cf. Experiment 1) in Experiment 3's analyses for a counterbalanced design, because this experiment's items were not assigned to participant lists to form consistent item groups (i.e. so that items could be put into groups would be seen in the same condition for any given participant list). One critical point made in Raaijmakers et al. (1999) is that it is appropriate, for designs in which matched experimental materials are used on all conditions and in the case of missing data (which we have created through our data removal procedure), to analyze the experimental data with by-participant ANOVAs only (i.e. calculating F_1 is sufficient). As was discussed in Section 2.1.1, we take our experimental materials to be appropriately matched to adhere to these requirements. So, the visual attention as well as lexical decision data produced in Experiment 3 were analyzed using repeated-measures ANOVAs on the by-participant-and-condition averages of the respective measures, with our experimental manipulations as a within-subject factors and participant list as a between-subject factor.

2.3.2 Results

The results section of this experiment is structured similarly to Experiment 1's results section. Please also refer to the 'Analysis' Section of Experiment 1 as well as to the Glossary for the meaning of specific analysis-related terms that are used in the following results section.

Deployment of visual attention

Time course of visual attention: overview The most obvious observation to be made on the time course of visual attention deployment (cf. Figures 2.17 and 2.18) is that the visual cue (the highlighting circle) in this experiment caused a (roughly equally) strong tendency to look at the highlighted object in all conditions within about 150 msec after the onset of the spoken word, corresponding to 200 msec after the onset of the highlighting cue and thus the time necessary to trigger a saccade. The probability of looking at the highlighted object peaked between 300 and 400 msec after word onset, reaching well above .8 in all conditions. In addition to that, the linguistic cue (the spoken word) caused looks to the critical object independently from the highlighting manipulation: One can see a clear tendency to look at the referent critical object when the baseline object is highlighted, starting from the end of the spoken word (cf. Figure 2.18 B)). The same, though weaker, pattern could be observed when the critical object was semantically related to the spoken word (cf. Figure 2.18 F)). At word onset, there were slight imbalances between critical object or baseline objects and the distractor on four conditions: when the critical object was a referent or the unrelated object and was highlighted, and when the critical object was phonologically related or semantically related and the baseline object was highlighted (cf. Figures 2.17 and 2.18 A), G), D), F)).

Assessing the visual saliency of the displayed objects: Highlighted- and critical-object inspection advantages before word onset Highlighted-object inspection advantages for the shifted time region before word onset (cf. Figure 2.19) were compared to zero using one-sample t-tests for each condition (cf. Table 2.1). In addition to that (as critical object and highlighted object were the same in those conditions in which the critical object is highlighted) the same tests were computed on critical-object advantages for those conditions in which the

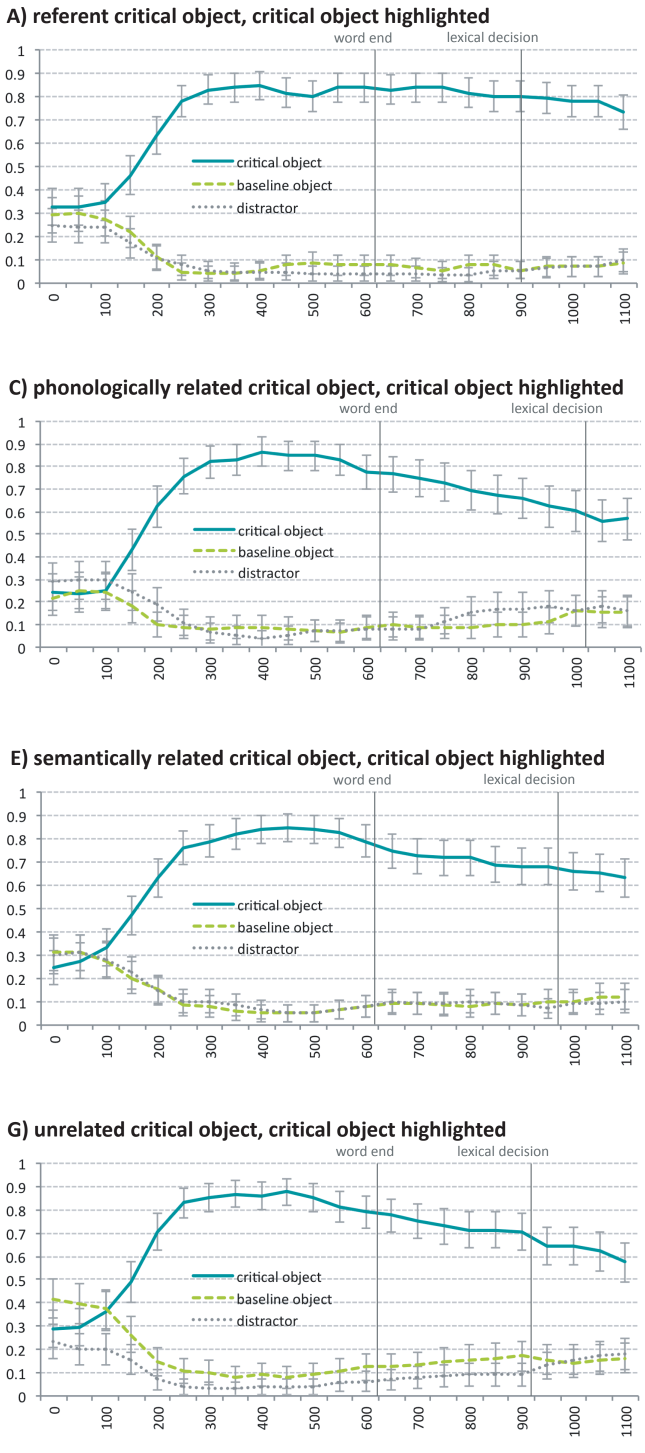

Figure 2.17: On this and the following page: Experiment 3: Critical-object, baseline-object and distractor inspection probabilities for 50-msec long time slots relative to the onset of the spoken word, with one graph (A to H) per condition. The time identifying a time slot denotes its end point. Error bars indicate 95% confidence intervals.

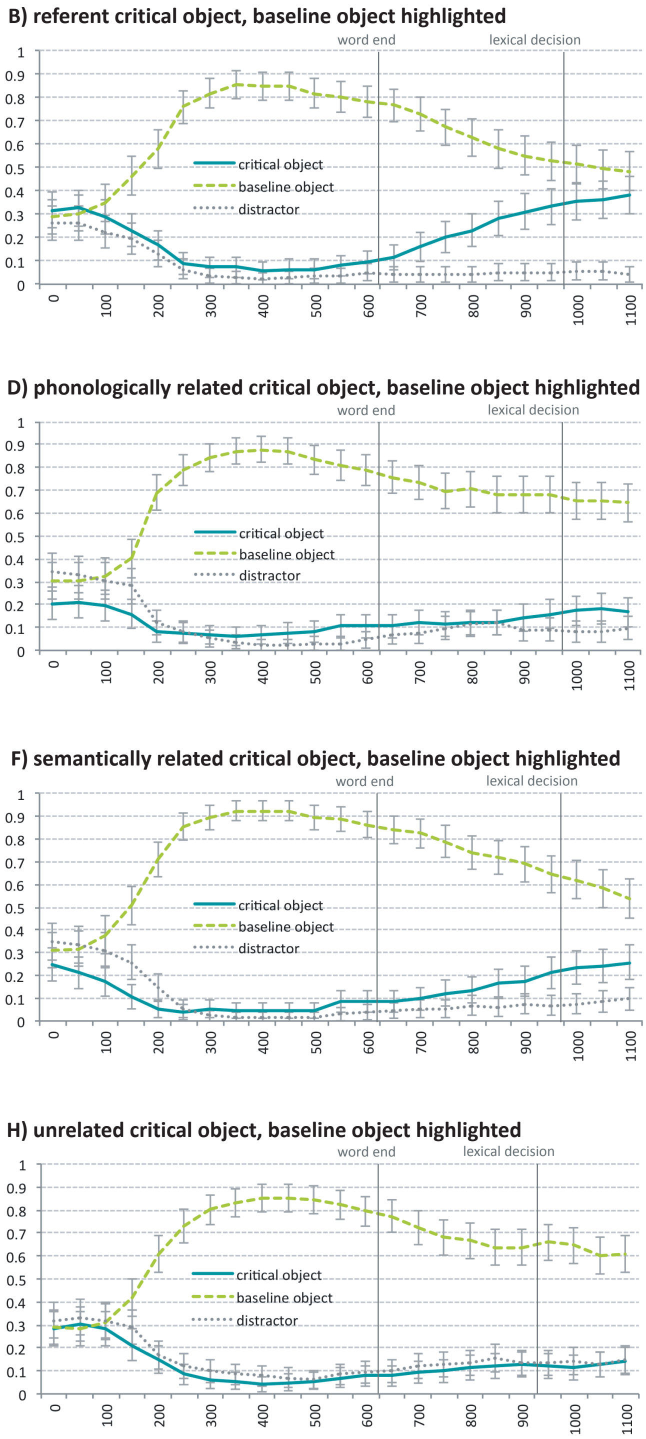

Figure 2.18: On this and the previous page: Experiment 3: Critical-object, baseline-object and distractor inspection probabilities for 50-msec long time slots relative to the onset of the spoken word, with one graph (A to H) per condition. The time identifying a time slot denotes its end point. Error bars indicate 95% confidence intervals.

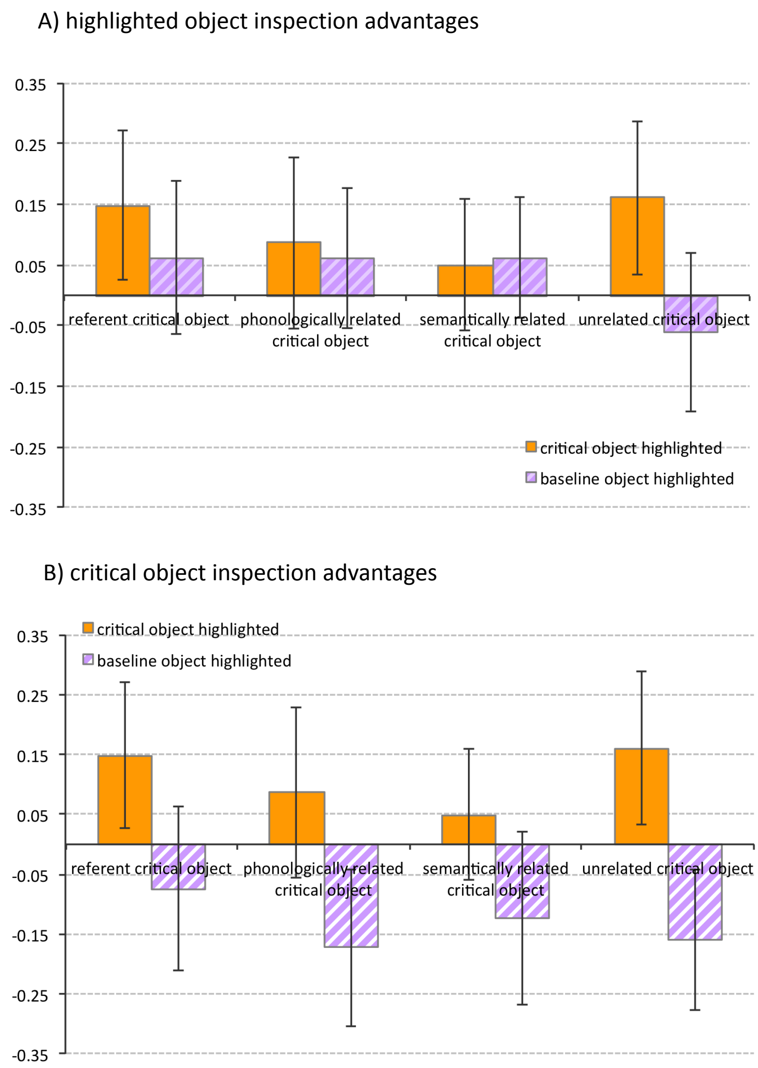

Figure 2.19: Experiment 3: Average highlighted-object (graph A)) and critical-object (graph B)) inspection advantages for the shifted time region before word onset, one bar per condition. Error bars indicate 95% confidence intervals.

baseline object was highlighted (cf. Table 2.2). Significant inspection imbalances were found on only two conditions: Before the onset of the spoken word, there was a highlighted-object advantage when the referent critical object was highlighted and a distractor advantage over the critical object when the critical object was phonologically related to the spoken word and the baseline object was highlighted.

Table 2.1: Experiment 3: Results of t-tests establishing whether highlighted-object inspection advantages before word onset are significantly different from zero.

condition	t-test by subjects	t-test by items
referent critical object, critical object highlighted	$t_1(46) = 2.3^*$ $p < .05$	$t_2(23) = 2.1^*$, $p < .05$
referent critical object, baseline object highlighted	$t_1(45) = .6$, $p = .54$	$t_2(23) = .6$, $p = .56$
phonologically related critical object, critical object highlighted	$t_1(36) = .1$, $p = .91$	$t_2(20) = .4$, $p = .66$
phonologically related critical object, baseline object highlighted	$t_1(43) = .43$, $p = .67$	$t_2(23) = .3$, $p = .75$
semantically related critical object, critical object highlighted	$t_1(45) = 1.0$, $p = .33$	$t_2(23) = .9$, $p = .36$
semantically related critical object, baseline object highlighted	$t_1(43) = .7$, $p = .47$	$t_2(23) = .9$, $p = .38$
unrelated critical object, critical object highlighted	$t_1(42) = .4$, $p = .66$	$t_2(23) = .3$, $p = .75$
unrelated critical object, baseline object highlighted	$t_1(46) = 1.6$, $p = .12$	$t_2(23) = 1.5$, $p = .15$

Table 2.2: Experiment 3: Results of t-tests establishing whether critical-object inspection advantages before word onset are significantly different from zero for those conditions on which the critical object was not highlighted.

condition	t-test by subjects	t-test by items
referent critical object, baseline object highlighted	$t_1(45) = 1.1$, $p = .23$	$t_2(23) = 1.2$, $p = .25$
phonologically related critical object, baseline object highlighted	$t_1(43) = 2.2^*$, $p < .05$	$t_2(23) = 2.1^*$, $p < .05$
semantically related critical object, baseline object highlighted	$t_1(43) = 1.3$, $p = .21$	$t_2(23) = 1.4$, $p = .17$
unrelated critical object, baseline object highlighted	$t_1(46) = 1.9$, $p = .01$	$t_2(23) = 1.7$, $p = .10$

Assessing the amount of visual attention shifted towards the highlighted and critical objects in response to the highlighting cue and the spoken word: Highlighted- and critical-object advantages for inspections started during the spoken word One-sample t-tests on highlighted-object advantages for the probabilities of inspections having been started during the shifted region from word onset to word end (cf. Figure 2.20 A)) showed signif-

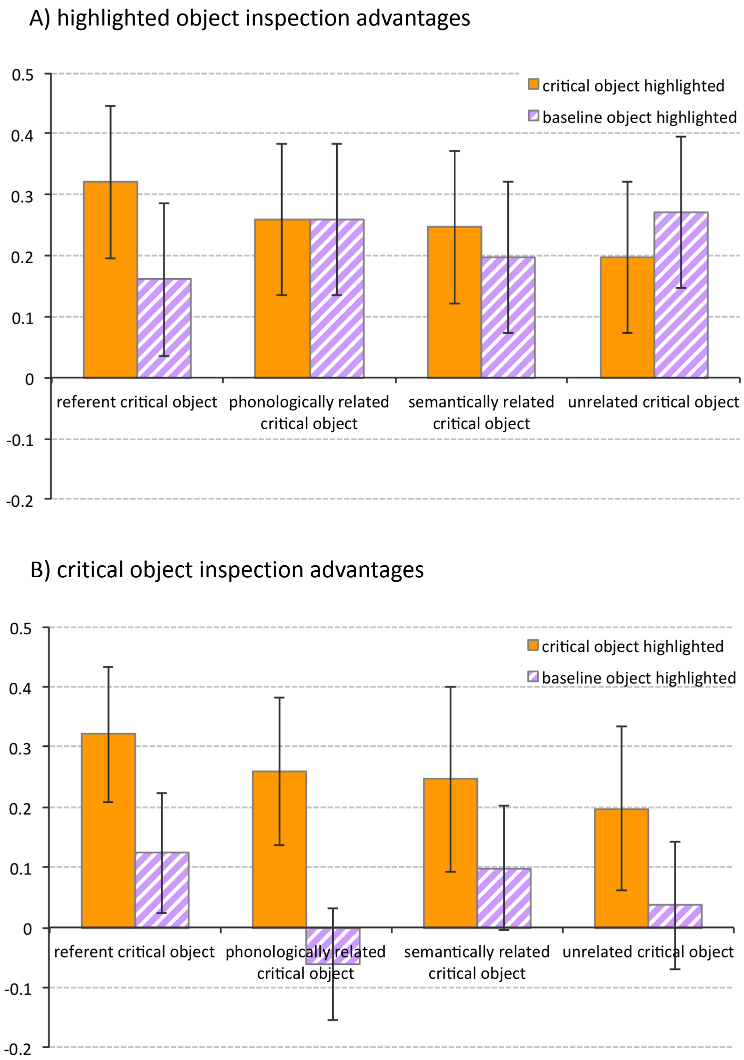

Figure 2.20: Experiment 3: Average highlighted-object (graph A)) and critical-object (graph B)) advantages in probabilities of inspections having been started during the shifted word region, one bar per condition. Error bars indicate 95% confidence intervals (calculated following Masson and Loftus (2003) for graph B)).

icant highlighted-object advantages for all conditions, confirming significant tendencies to shift gaze towards the highlighted object during the spoken word for every condition (cf. Table 2.3). An ANOVA on this data showed no significant effects of either of the two experimental factors (critical object relatedness: $F(3,60) = 1.7$, $p = .17$; highlighted object: $F(1,20) = 1.4$, $p = .26$) or an interaction ($F(3,60) = .9$, $p = .45$). Note: the lack of an effect of the highlighted-object factor on highlighted-object inspection advantages during the spoken word is to be expected, as the highlighted object, per definition, was highlighted on every condition. Thus, if participants followed the highlighting cue reliably, highlighted-object inspection advantages should be equally high on all conditions and not be affected by any of the experimental factors.

Table 2.3: Experiment 3: Results of one-sided t-tests establishing whether highlighted-object inspection advantages during the shifted time region of the spoken noun were significantly larger than zero.

condition	**t-test by subjects**	**t-test by items**
referent critical object, critical object highlighted	$t_1(46) = 7.3$***, p <.001	$t_2(23) = 8.4$***, p <.001
referent critical object, baseline object highlighted	$t_1(45) = 5.7$***, p <.001	$t_2(23) = 6.4$***, p <.001
phonologically related critical object, critical object highlighted	$t_1(36) = 5.0$***, p <.001	$t_2(20) = 5.1$***, p <.001
phonologically related critical object, baseline object highlighted	$t_1(43) = 4.9$***, p <.001	$t_2(23) = 4.6$***, p <.001
semantically related critical object, critical object highlighted	$t_1(45) = 4.9$***, p <.001	$t_2(23) = 6.8$***, p <.001
semantically related critical object, baseline object highlighted	$t_1(43) = 4.1$***, p <.001	$t_2(23) = 5.4$***, p <.001
unrelated critical object, critical object highlighted	$t_1(42) = 1.7$***, p <.001	$t_2(23) = 5.6$***, p <.001
unrelated critical object, baseline object highlighted	$t_1(46) = 6.0$***, p <.001	$t_2(23) = 6.2$***, p <.001

Shifts of visual attention to the critical object in response to the spoken word (cf. Figure 2.20 B)) were assessed by analyzing critical-object advantages for the probabilities of inspections having been started during the shifted region from word onset to word end, only for those conditions on which the baseline object was highlighted (as critical-object inspection advantages overlapped with highlighted-object advantages for those conditions on which the critical object was highlighted and were thus already covered by the analyses of highlighted object advantages above). One-sample t-tests showed significant critical-object advantages for the non-highlighted referent critical object as well as the semantically related critical object. Non-highlighted phonologically related critical objects were not looked at in response to the spoken word.

Table 2.4: Experiment 3: Results of t-tests establishing whether critical-object inspection advantages during the shifted time region of the spoken word were significantly different from zero for those conditions on which the critical object was not highlighted.

condition	**t-test by subjects**	**t-test by items**
referent critical object, baseline object highlighted	$t_1(45) = 3.7^{***}, p < .001$	$t_2(23) = 3.5^{**}, p < .01$
phonologically related critical object, baseline object highlighted	$t_1(43) = .4, p = .71$	$t_2(23) = .3, p = .78$
semantically related critical object, distractor highlighted	$t_1(43) = 1.9, p = .06$	$t_2(23) = 2.6^{*}, p < .05$
unrelated critical object, distractor highlighted	$t_1(46) = .5, p = .59$	$t_2(23) = .6, p = .53$

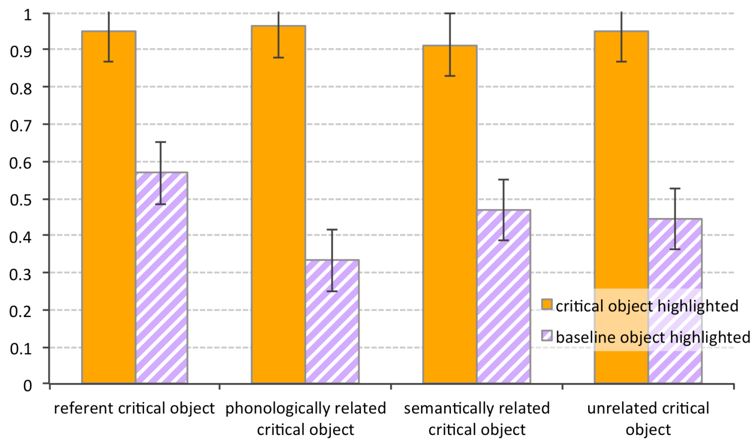

Figure 2.21: Experiment 3: Average critical-object inspection probabilities for inspections having lasted during the second before lexical decision, one bar per condition. Error bars indicate 95% confidence intervals calculated following Masson and Loftus (2003).

Assessing the potential of highlighted and critical objects to affect word comprehension: Highlighted- and critical-object inspection probabilities prior to lexical decision

The probability to have looked at the highlighted object at least once during the second before the lexical decision response was around .94 in all conditions and was not significantly affected by either critical-object relatedness ($F(1.9,38.4) = 1.0$, $p = .36$) or highlighted object ($F(1,20) = 0.5$, $p = .5$, interaction: $F(3,60) = 1.5$, $p = .82$).

The probability of inspecting the critical object at least once during the second before the lexical decision (cf. Figure 2.21) was, as expected, strongly affected by the highlighted-object factor ($F(1,20) = 171.2$, $p < .001$), such that the critical object was looked at more when it was highlighted than when it was not (all $p < .001$ in pairwise comparisons). Critical-object relatedness had no significant effect ($F(3,60) = 2.2$, $p = .10$). There was, however, a significant interaction between critical-object relatedness and highlighted object ($F(3,60) = 2.9$, $p < .05$), caused by significantly more inspections to the non-highlighted referent critical object than to the non-highlighted phonologically related critical object, i.e. an effect of the critical object's relatedness when the baseline object was highlighted, but not when the critical object was highlighted.

Speed and accuracy of lexical decision

An ANOVA on the average number of accurate trials (cf. Figure 2.22) revealed a significant main effect of critical-object relatedness ($F(2.1, 85.2) = 5.9$, $p < .01$) and a significant interaction of critical-object relatedness and highlighted object ($F(2.1, 82.8) = 3.8$, $p < .05$). The critical-object relatedness factor had no significant effect on response accuracy ($F(1,40) = 3.4$, $p = .07$). Pairwise comparisons showed response accuracy to be lower when the phonologically related critical object was highlighted than when any other critical object was highlighted (all p < .05).

Lexical decision latencies (cf. Figure 2.23) were affected significantly only by the relatedness of the critical object ($F(3,60) = 4.2$, $p < .01$). This effect was carried by, when the critical object was highlighted, the referent and unrelated critical objects coinciding with significantly faster lexical decision latencies than phonologically or semantically related critical objects (all $p < .05$ in pairwise comparisons). In addition to this, when the critical object was the spoken referent, the critical object being highlighted coincided with faster lexical decisions than the baseline object being highlighted ($p < .05$ in pairwise comparisons). The highlighted object factor did not affect lexical decision latencies ($F(1,20) = 2.1$, $p = .16$), nor did the interaction between the highlighting and relatedness factors ($F(3,60) = 2$, $p = .12$).

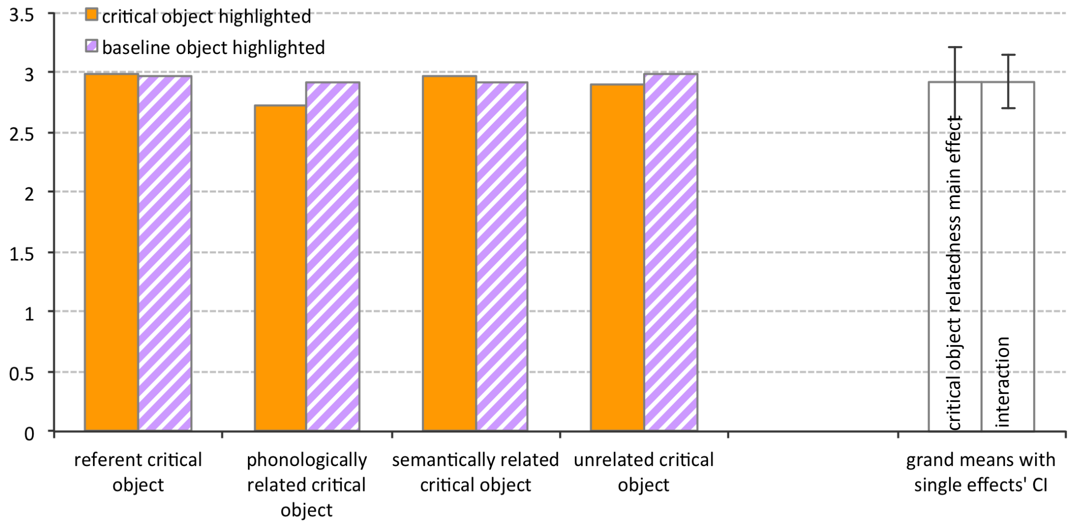

Figure 2.22: Experiment 3: Average number of correct trials per participant per condition. Error bars indicating a 99% confidence interval for the critical-object relatedness main effect and a 95% confidence interval for the interaction are plotted on separate grand mean bars following Masson and Loftus (2003).

2.3.3 Discussion

In Experiment 3, our analyses showed the following patterns of visual attention deployment: Most importantly, participants followed the highlighting cue reliably, resulting in a large likelihood to inspect the highlighted object during the spoken word on all conditions. There were slight imbalances between inspection probabilities for the different displayed objects at word onset, but later shifts of attention to the highlighted object were equally likely and well-pronounced on all conditions, suggesting that these early imbalances can be ignored for the purpose of the interpretation of this experiment. In addition to that, there was never the same pattern of imbalances on two conditions on which the same objects were displayed (i.e. condition pairs with the same critical object type), so it can be concluded that these early imbalances have to be attributed to chance rather than the displayed objects' visual salience. There is also a tendency to shift gaze towards the critical object in response to the spoken word on those conditions on which the baseline object is highlighted and the critical object is a referent or a semantically related object. This suggests that participants did recognize the displayed objects sufficiently during the preview time that they could launch NCEs to the related critical objects in response

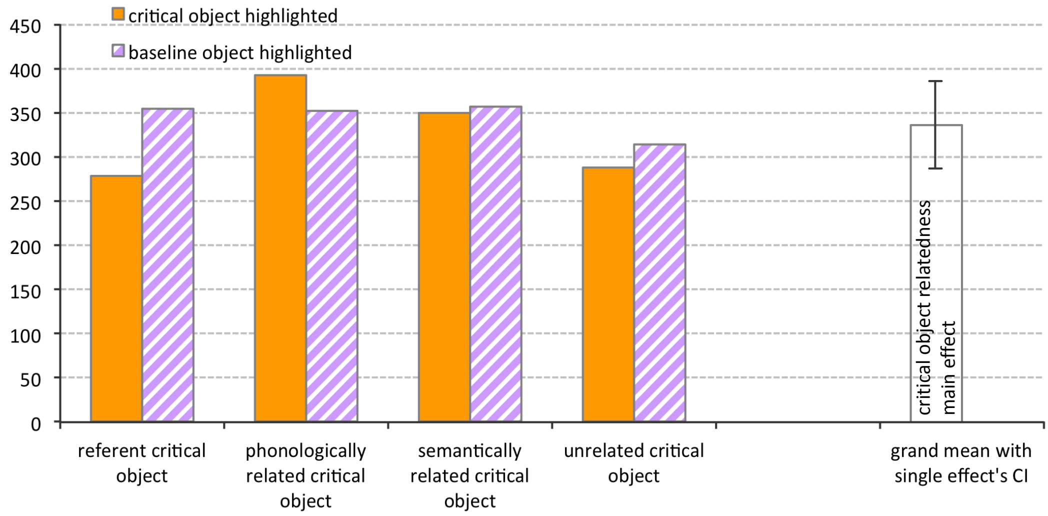

Figure 2.23: Experiment 3: Average lexical decision latencies in msec measured from word end for the entire data set, one bar per condition. Error bars indicating a 99% confidence interval for the critical-object relatedness main effect are shown on a separate grand mean bar following Masson and Loftus (2003)

.

to the spoken word, although the visual cue at word onset caused them to look at the highlighted (unrelated) object first. Why then did they not look at the phonologically related critical object when the baseline object was highlighted? Experiment 2 (Figure 2.11) shows that the tendency to look at the phonologically related critical object in response to the spoken word occurs right after word onset and declines from about word end. It is likely that, when participants followed the visual cue to look at the highlighted baseline object right after word onset, the portion of the spoken word that overlapped with the phonologically related critical object and could trigger looks to it was already over by the time at which participants had finished inspecting the highlighted object.

In order to discuss this experiment's lexical decision results, it is important to first establish our baseline condition. Ideally, the two conditions on which the critical object was unrelated to the spoken word should have shown the same lexical decision latencies (as the highlighted object, both when it was the unrelated critical object as well as when it was the baseline object, was unrelated to the spoken word and the displayed objects were the same on both conditions) and should both have served as baseline conditions for the lexical decision analyses. Unexpectedly, however, lexical decision latencies when the unrelated critical object was highlighted were not slower, but rather as fast as when the referent critical object was highlighted. Experiment 4 will show that this is not due to the unrelated object having a beneficial relation to the spoken word, so we are led to believe that this effect is spurious. So, for this discussion, the condition on which the critical object was the unrelated object and the baseline object was highlighted will serve as the lexical decision baseline condition. Equal lexical decision latencies on this and all the other baseline object highlighting conditions support this definition.

The most clear result from this experiment's lexical decision latency analyses is, again, that looking at a spoken word's referent while comprehending its name speeds up lexical decision to this name. Crucially, there is no such benefit for the speed of lexical decision when the referent is present but not looked at during word comprehension: when participants had to look at the unrelated baseline object while comprehending the spoken word, even though the referent had been seen before word onset and looked at to a certain degree between word end and the lexical decision response, lexical decision latencies were the same as when the referent had never even been displayed. There are two possible explanations for this: The first one is that the presence of a visual referent can only benefit word comprehension when the referent is looked at during the life time of the spoken word. A second potential explanation would be that the interference from looking at the baseline object and the benefit from looking at the referent before and/or after the spoken word canceled each other out to yield lexical decision latencies matching those on conditions when the referent had never been present. The current results do not speak for or against any of these two accounts, but Experiment 7 will address this question again.

Looking at the phonologically related critical object during word comprehension did not affect lexical decision latencies. It did, however, cause more lexical decision errors than on any other condition. Error trials were removed from the data set before running the lexical decision latency analyses, which further reduced the already low number of trials on this condition and may have led to lexical decision latencies not being significantly (but still numerically) larger on this than on other conditions. Having to look at the baseline object in the presence of the phonologically related critical object did not interfere with lexical decision latencies or accuracy. Our first three experiments' results thus remain inconclusive with regard to the interaction between the processing of phonologically related visual objects and noun comprehension.

The semantically related object's presence did not affect lexical decision latencies, whether it was looked at during the spoken word's lifetime or not. This was the case even though shifts of attention towards the semantically related critical object were made independently of the highlighting manipulation, which shows that its relation to the spoken word was recognized by participants. This conforms with the results of Experiments 1 and 2.

In sum, this experiment showed a solid word comprehension benefit from making eye movements to a spoken word's referent while comprehending that spoken word. Having to look at a phonologically related object interfered with word comprehension (causing lexical decision errors instead of slowing down correct lexical decisions). Crucially, both these effects were eradicated when the respective related objects were present in the display but an unrelated object was looked at during comprehension. Thus, we can conclude that eye movements to objects have an effect on word comprehension. Looks to a spoken word's referent speed up that word's comprehension, while looks to a phonologically related object make it less accurate. Both findings would fit an account in which noun-contingent eye movements are made *in order to* benefit from them, as looks to phonologically related objects are triggered while the spoken input still matches the displayed object (during the overlapping phonemes of word and object). The next section will further explore the mechanisms by which noun-contingent referent looks can benefit word comprehension. The question of why we continue to find no effect of a semantically related object on word comprehension is addressed in Experiment 4.

2.4 Lexical decision with single visual objects: Experiment 4

This thesis' first three experiments addressed the issue of whether presenting our visual critical objects simultaneously with the spoken word would actually reliably produce lexical decision effects similar to previous studies' priming effects. Given the somewhat unexpected findings in our paradigm, which combines scenes containing multiple objects with a lexical decision task, Experiment 4 was run as a simple priming study. This way we could determine whether or not we could replicate previous findings with our stimuli.

Participants only saw one single visual object before and during comprehension of the spoken word that was the subject of the lexical decision task (cf. Figure 2.24). No eye-tracking data was recorded. On experimental trials, the displayed object was either one of the four critical object types from the previous three experiments (referent, phonologically related, semantically related or unrelated) or the unrelated baseline object used in Experiment 3. This last type of visual object was included in order to confirm that the benefit from looking at the unrelated critical object compared to looking at the unrelated baseline object, observed in Experiment 3, had simply been an artefact of that study. Based on previous findings discussed in Section 1.3, replication should lead to lexical decision benefits from seeing the referent or the semantically related object and interference from seeing the phonologically related object, all when seeing the unrelated critical object. Also, there should be no difference in lexical decision latencies between the two conditions with unrelated visual objects (unrelated critical object and baseline

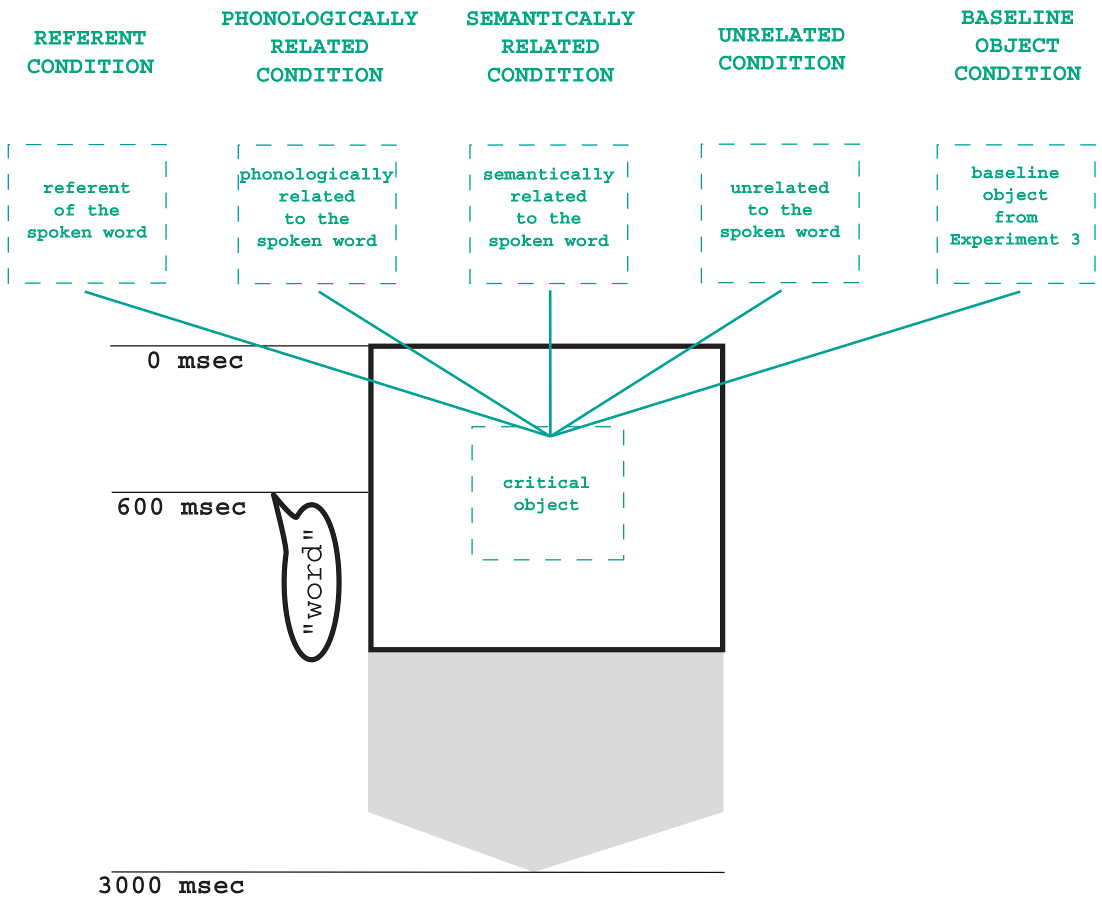

Figure 2.24: A sample trial of Experiment 4. The display, containing only one object (the spoken word's referent / a phonologically related object / a semantically related object / an unrelated critical object / an unrelated baseline object), was shown with a preview of 600 msec. Then, the spoken word was played while the display remained unchanged. Participants gave their lexical decision response via a button press. Every trial had an overall duration of three seconds.

object from Experiment 3).

2.4.1 Method

Participants

Thirty-six native German speakers from the Saarland University community took part in this experiment for a payment of five Euros. Six participants' data was not included in the analyses because participants turned out to be bilingual, were overall unable to accomplish the task or had already taken part in an earlier experiment that used similar stimuli, or because construction works disrupted the experiment. Overall, the data of 30 participants was analyzed.

Materials

This experiment's materials were taken from the 28 experimental items of Experiment 2. The display now only contained a single object, which could be the spoken word's referent, a phonologically related object, a semantically related object, an unrelated object (these four were the critical objects in Experiment 2) or an unrelated baseline object (this was the unrelated baseline object from Experiment 3. For a sample set of stimuli, see Figure 2.25)[30] Despite its name, the unrelated baseline object did not serve as a lexical decision baseline in the current experiment - this function was taken by the unrelated object (the one that had been a critical object in Experiment 2).

The visual object was displayed in the centre of the screen, where it filled a 200 x 200 px square and thus took up 6 °of the participant's field of view. The spoken stimuli were the same as in Experiment 2, with an average duration of 621 msec across all experimental trials.

In addition to the 28 experimental items, there were 56 filler items: 14 word items and 42 non-word items, on six of which the spoken non-word had the same onset as the displayed object's name. Also, ten practice trials resembled all types of trials occurring in the experiment itself.

Design This experiment had a 5 (displayed-object relatedness) x 1 within-subject design, with five subject lists.

Procedure

This experiment was a reaction time experiment, conducted on a PC with a 19" display, using Experiment Builder software and Cedrus response pad (as in the previous studies).

[30]Note that the current experiment had 28 experimental items while Experiment 3 only had 24.

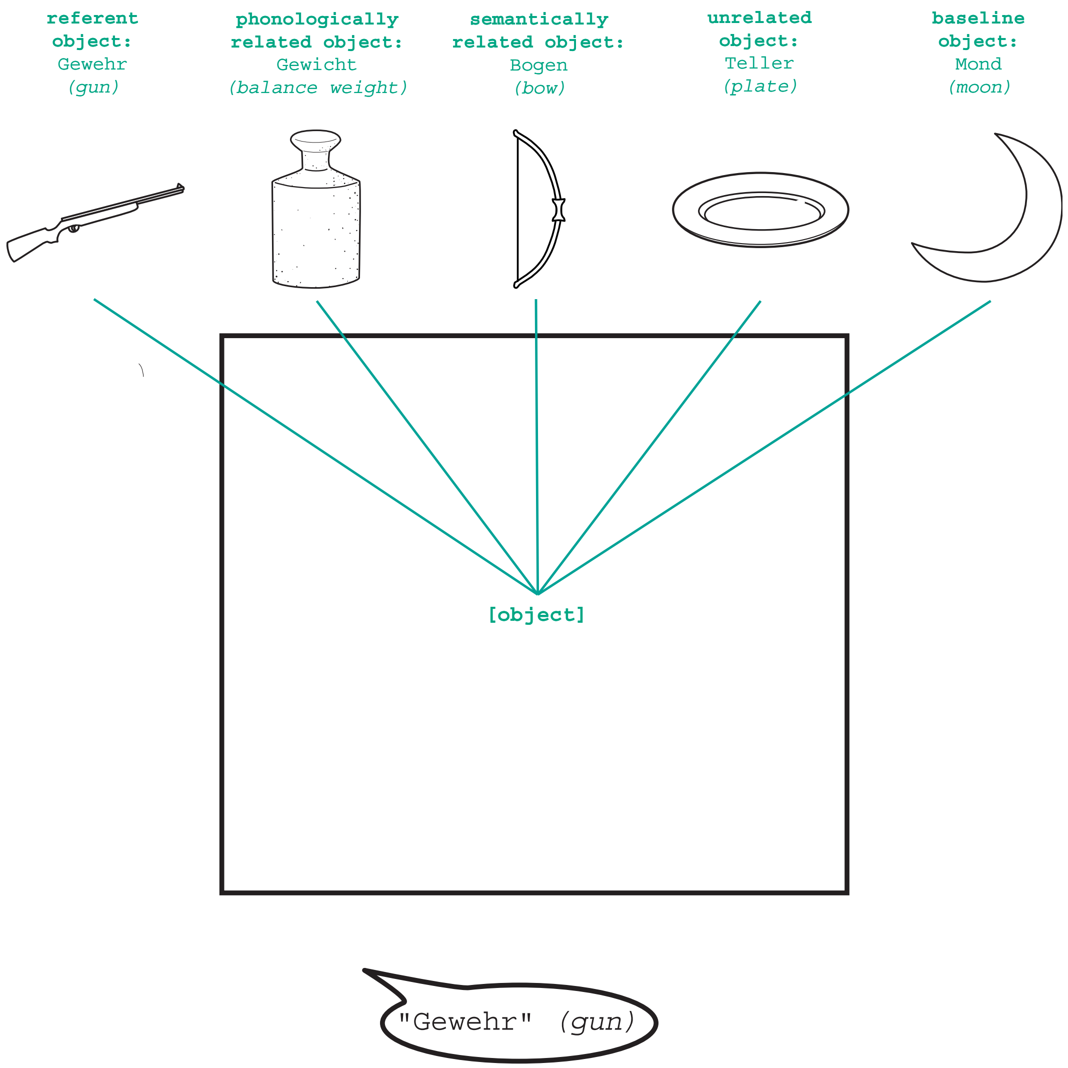

Figure 2.25: The materials for one of Experiment 4's items. Participants were presented with a display a single object, whose identity changed depending on the condition: it could either be the spoken noun's referent, a phonologically related object, a semantically related object, an unrelated object or the unrelated baseline object from Experiment 3.

Participants were instructed to make their lexical decision as quickly and as accurately as possible. They were asked to look at the displayed objects, but to base their word vs. non-word decision only on the spoken stimuli.

Every trial began with a display preview of 600 msec before word onset. The duration of this preview was chosen to resemble the time participants spent gazing at the critical object before word onset in Experiment 3, which was on average 566 msec. The object remained on the display for the entire three-second duration of a trial.

The order of trials was pseudo-random and different for each participant, with each experimental item being preceded by two filler items, the second one of which was always a non-word trial. Due to an error in the implementation of the experiment's script, only the first 72 trials of each participant's trial list were actually run, which led to the loss of four random experimental trials per subject.

Analysis

Data removal 3.5% of the trials from the initial data set were removed because of incorrect or premature lexical decision responses. In addition to that, three lexical decision latencies were considered outliers because their distance from their means exceeded two standard deviations and one additional latency was removed based on the inspection of scatterplots (it was larger than 1000 msec). The resulting data set contained 95.9% of the initial data set's data points.

Inferential statistics This experiment, as the previous one, was (unintentionally) not designed to have consistent item groups, so we again conducted simple ANOVAs with the displayed object's relatedness as a within-subject factor and participant list as a between-subjects factor.

2.4.2 Results

Speed and accuracy of lexical decision

Participants' accuracy in performing their lexical decision task was on average 92% (4.6 correct responses out of 5 trials per participant and condition) and was not affected by the type of the displayed object ($F(4,104) = .5$, $p = .77$) .

Lexical decision latencies (cf. Figure 2.26) were affected significantly by the type of the displayed object ($F(4,104) = 24.6$, $p < .001$). Pairwise comparisons showed the following significant differences between conditions: Participants were significantly faster to make their lexical decision when the displayed object was the referent than when it was any other object (all $p < .01$)

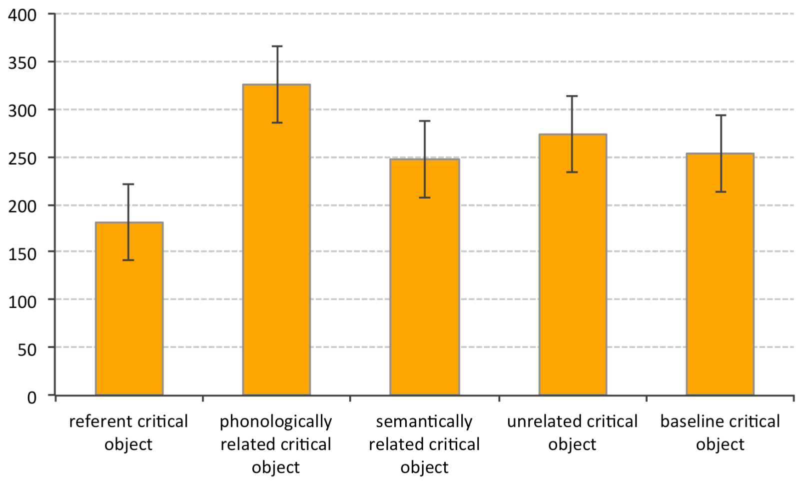

Figure 2.26: Experiment 4: Average lexical decision latencies measured from word end, one bar per condition. Error bars indicate 99.9% confidence intervals calculated following Masson and Loftus (2003).

and significantly slower when the displayed object was the phonologically related critical object than on any other condition (all $p < .01$). So, relative to our baseline condition (when the displayed object was the unrelated critical object), we found facilitation for the referent condition, interference for the phonologically related condition, and neither facilitation nor interference for the semantic relatedness and unrelated baseline object (from Experiment 3) conditions.

2.4.3 Discussion

In this priming study we could confirm the solidly established benefit from the presence of a referent object during the comprehension of a spoken word on lexical decision latencies to that word. Also, the interference from seeing a phonologically related single object, observed in Experiment 1, could be replicated. In addition, we established that the benefit from looking at the unrelated critical object in Experiment 3 must have been spurious, as it could not be confirmed in this experiment: lexical decision latencies when faced with that unrelated critical object vs. the unrelated baseline object were the same. Crucially, we did not observe any facilitatory effect of the presence of a semantically related visual object on the comprehension of a simultaneously perceived noun, confirming Experiment 1 - 3's findings. Even though we had expected this semantic facilitation when designing the first three experiments of this thesis based on empirical evidence, our result does not actually stand in contrast to at least two previous priming studies: Levelt et al. (1991) (Experiment 3) as well as Anderson and Holcomb (1995) (Experiment 2) similarly find no semantic facilitation when prime and target overlap temporally in a lexical decision task. Also, McQueen and Huettig (2005)'s semantic facilitation effect, in a study very similar to our Experiment 4, is small and potentially non-significant (there are no claims about significance in McQueen and Huettig (2005)).

2.5 Summary

The first four experiments of this thesis provide clear evidence that noun-contingent eye movements can affect noun comprehension. Experiments 1 and 2 show that a related object image affects the speed of lexical decision when there are several objects visible at once and participants can choose whether or not to look at the related object. Referent objects were looked at reliably in response to a spoken noun, and lexical decision latencies were faster when a referent was in the display than when other related or only unrelated objects were shown. When, in Experiment 3, visual attention was directed towards the spoken word's referent during the word's comprehension, comprehension benefited compared to when visual attention was directed toward an unrelated object. Specifically, comprehension benefited only when the spoken word's referent was looked at during comprehension, not when it was present in the display but not looked at upon word onset. As the deployment of visual attention during the spoken word was controlled experimentally via a visual cue, this means that the object that people look at while listening to a spoken noun does influence the speed with which they comprehend the noun. Results are consistent for referent objects, but less clear for objects phonologically related to the spoken word. Experiment 4 shows that our phonologically related objects do interfere with an auditory lexical decision task when only one such object is in the display. Experiments 1 and 2 provide conflicting evidence about the correlation between inspection probabilities and lexical decision interference for phonologically related objects, while Experiment 3 adds evidence for looks to a phonologically related object negatively affecting simultaneous noun comprehension. In sum, studies 1 to 4 support the notion that the object one fixates while comprehending a spoken noun will influence the speed of lexical access for that noun.

The following three experiments will concentrate on the interaction between a spoken noun and its depicted referent. They build on the result that the comprehension of a spoken noun profits from the simultaneous visual perception of its referent. In the next chapters, Experiments 5 to 7 further investigate the mechanisms that might lie behind a beneficial effect of noun-contingent eye movements on noun comprehension.

3 Priming mechanisms as the source of a referent-look comprehension benefit

Experiment 3 revealed that participants who see a spoken word's referent before that word's onset profit from looking at the referent during word comprehension, compared to looking at an unrelated object during the lifetime of the spoken noun. This benefit was observed only as a result of referent looks triggered at word onset - if the referent could not be looked at during the spoken word, even its presence before word onset had no positive effect on lexical decision latencies at all. Considerations in the priming and lexical decision frameworks offer up a number of hypotheses about how referent NCEs could benefit noun comprehension (cf. Section 1.5): Firstly, the processing of the referent's visual image could benefit the processing of the noun, either as a priming effect from viewing the referent during the *preview period* or as a immediate priming from seeing it during *comprehension*. The third hypothesis states that seeing the non-matching visual image of an *unrelated object* could have a detrimental effect on comprehension. This chapter reports our findings about whether one or more of these factors are responsible for an effect of referent looks on noun comprehension.

3.1 Priming vs. avoidance of interference: Experiment 5

Experiment 5 was designed so that it would tease apart the explanations for the benefit of referent looks described above. During a preview period participants inspected the three displayed objects: a critical object, which could either be the spoken word's referent or an unrelated object, an unrelated baseline object and a distractor.[31] At word onset, a circle appeared for 100 msec

[31] At first glance, it might seem unnecessary to distinguish between three kinds of objects that are unrelated to the spoken word in this experiment: the unrelated critical object, the unrelated baseline object and the distractor. However, each one had its specific function. The *distractor* object was, as in the previous experiments, present on all conditions, always unrelated to the spoken word, never highlighted by the visual cue and thus served as a baseline object for visual attention analyses. The unrelated *baseline* object was also present on all conditions and unrelated to the spoken word, but it could be highlighted by the visual cue and served as a lexical decision baseline similar to the previous experiments. Indeed, one of the issues addressed in this experiment was whether such an unrelated object is a good lexical decision baseline or whether it itself would interact with lexical decision process. The unrelated *critical* object was necessary to replace the referent object on those conditions on which

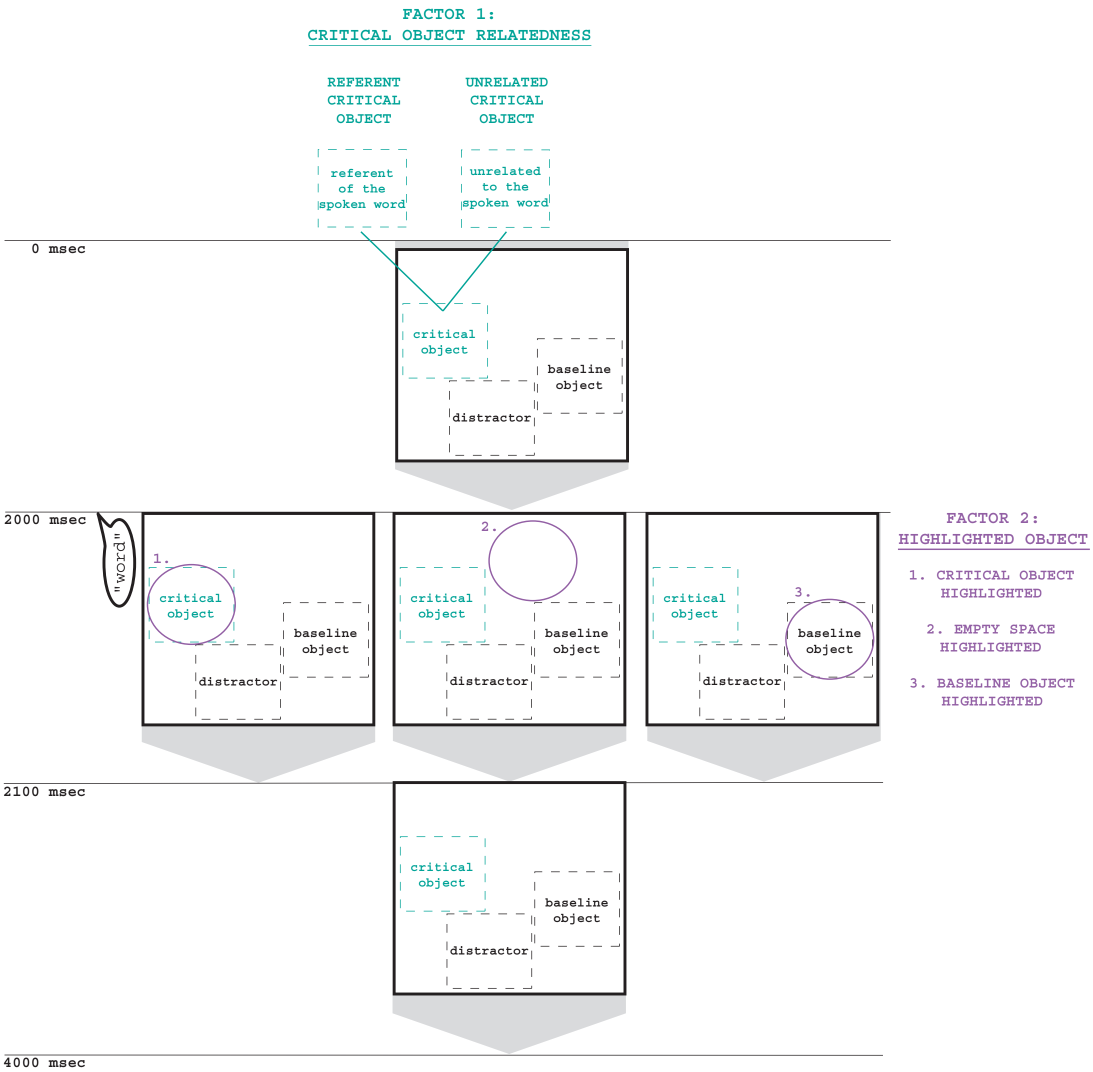

Figure 3.1: A sample trial of Experiment 5. The display, containing the critical object (the spoken word's referent / an unrelated object), an unrelated baseline object and one distractor, was shown with a preview of two seconds. Then, the spoken word was played. At the spoken word's onset, a circle appeared on the screen and remained there for the duration of 100 msec. This circle drew attention to the critical object, the baseline object, or an empty position on the screen. Participants gave their lexical decision response via a button press. Every trial had an total duration of four seconds.

around either the critical object, the baseline object or an empty display location (cf. Figure 3.1). Once again, participants were instructed to pay attention to the highlighted display region and make their lexical decision on the spoken stimulus as quickly and accurately as possible.

By manipulating the presence of a referent for the spoken word in the display as well as whether that referent (if present), an unrelated object or no object was fixated during word comprehension, we can explore the following specific issues:

A first possibility is that having identified the referent object during preview could prime the spoken word and could thus, in theory, speed up its comprehension no matter which objects are looked at during comprehension. Experiment 3's results (no difference in lexical decision latencies between when the referent has been seen during the preview but the baseline object was fixated during word comprehension and when the referent has never been seen) speak against this, but a comparison between lexical decision latencies on those conditions on which the referent is displayed but not looked at (i.e. when the critical object is the referent and the highlighted 'object' is empty space or the baseline object) and those conditions on which the referent has never been displayed (i.e. when the critical object is an unrelated object and the highlighted 'object' is empty space or the baseline object) of this experiment can be used to confirm our previous results. A variant of this hypothesis (cf. 'preservation of priming' in Section 1.5) states that any priming from the referent object can only affect the processing of the spoken word if there are no additional non-matching objects being processed visually during word comprehension. So, looks to either the spoken word's referent or empty space during word comprehension would permit the concept of the referent, as primed during the preview period, to affect comprehension beneficially. This hypothesis would be supported by faster lexical decisions when the critical object is the referent and the critical object or empty space are highlighted than when the baseline object is highlighted. When the critical object is unrelated to the spoken word, no priming from the referent could ensue and thus there should not be a difference between the different levels of the highlighting factor with an unrelated critical object.

Secondly, it can be hypothesized that noun-contingent eye movements to referents benefit noun comprehension by avoiding a potential interference that would arise from perceiving a non-matching visual object during comprehension (cf. 'avoidance of interference' in Section 1.5). Supportive evidence for this hypothesis would be that looking at empty space while perceiving the spoken word produces consistently faster lexical decision latencies than looking at unrelated objects. Crucially and in contrast to the 'preservation of priming' hypothesis above, this should hold whether the referent object is co-present in the display or not.

Finally, it is conceivable that it is just the access to the immediate visual image of an instance

there should not be a referent in the display.

of the concept of the spoken word that makes noun-contingent eye movements beneficial for word comprehension (cf. 'referent perception benefit' in Section 1.5). If this is the case, having to look at the referent object from word onset should produce faster lexical decision latencies than looking at empty space or the baseline object (when the referent is present, i.e. when the critical object is the referent).

3.1.1 Method

Participants

Fifty-six native German speakers with normal or corrected-to-normal vision took part in this experiment and were paid five Euros in compensation. The data of eight participants was not included in the analyses because they guessed correctly what the purpose of the experiment was, they followed the highlighting manipulation very poorly as assessed by the experimenter, they had obvious difficulties coping with the task or because of technical problems. Data from the remaining 48 participants contributed to the analyses.

Materials

For this experiment, a new set of 48 items were assembled from the objects used in the previous experiments. Each experimental item consisted of a spoken word and four visual objects: the spoken word's referent and an object unrelated to the spoken word (these were the two possible critical object types), an unrelated baseline object and a distractor (cf. Figure 3.2).

Referent, unrelated object, baseline object and distractor were selected to be unrelated in shape, semantics[32] and name onset. All objects used in this experiment had been recognized consistently in the naming pre-tests described with Experiment 1.[33] Name length was equal on average for the referent, unrelated and baseline objects across items. Distractors were slightly longer than the referent and the unrelated object by on average 0.15 syllables across items, with a maximum length difference of three syllables. Name frequency differences between the referent, unrelated and baseline object were on average zero frequency classes across items, with a maximum differences of six frequency classes in both directions, but the majority of differences amounting to two or less frequency classes (above 75% of items). The distractor's name was slightly more frequent than the referent's and unrelated object's (on average .35 frequency classes across items), with one single maximal difference of eight frequency classes.

[32] The same criteria as in Experiment 1 were applied to intuitively assess two objects' semantic relatedness.

[33] This means for a displayed object that at least 90% of naming pre-test participants named it with the intended or a synonymous name, and no more than 10% of them named it incorrectly.

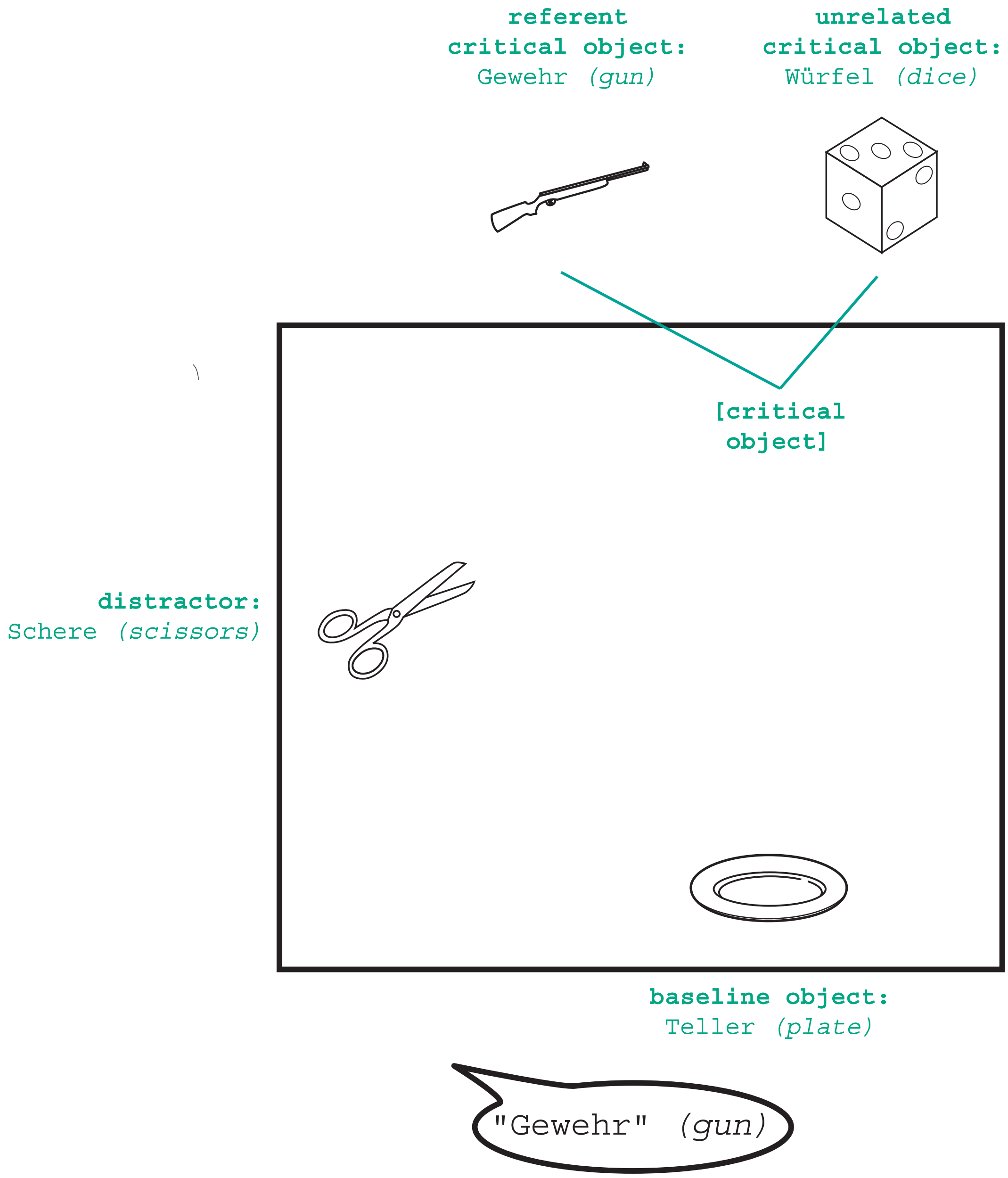

Figure 3.2: The materials for one of Experiment 5's items. Participants were presented with a display containing three inanimate objects: the critical object, an unrelated baseline object and a distractor. The critical object's identity was determined by the critical-object relatedness factor: it could either be the spoken noun's referent or an unrelated object. The display as it is shown here was presented from trial onset until trial end, with an additional circle appearing for 100 msec at word onset (cf. Figure 3.1).

Each display contained three objects that occupied three positions of a square layout: the baseline object, the distractor and the critical object. These were black-and-white line drawings that filled a 130 x 130 px square each and covered 4 °of participants' visual angle. The highlighting cue was a purple circle with a diameter of 200 px that could surround either the critical object, the baseline object, or the fourth, empty, grid position.

For each item, the layout that determined object positions in the display had one out of six possible rotations so as not to invite initial scanpaths akin to top-left-to-bottom-right reading. Each item had one fixed layout rotation and one fixed assignment of the three displayed objects to positions, so that layout rotations and object positions were balanced across items.

The spoken stimuli were recorded by a female native German speaker. The experimental items' spoken word had an average duration of 818 msec.

There were 48 experimental trials and 48 filler trials in this experiment. All filler trials had spoken non-words instead of spoken words and otherwise resembled the experimental trials: The empty space was highlighted on one third of the filler trials, objects in all possible screen locations were highlighted equally often and all six grid rotations were used equally often as well. There was never any phonological overlap between the spoken stimulus and the displayed objects on filler trials (nor was there on experimental trials, for that matter). In addition to these, twelve practice trials resembled all the item types used in the experiment.

Experiment 5 had a 2 (critical-object relatedness) x 3 (highlighted object) within-subject design.

Procedure

The equipment used in this experiment was the same as in the previous experiments: eye movements were measured with an EyeLink I system and lexical decision latencies recorded using a Cedrus response pad.

Participants received the written instruction to make their word - non-word decision about the utterance they heard as quickly and accurately as possible, to inspect the displayed images at the beginning of each trial and to pay special attention to the region highlighted by the circle. These points were repeated orally by the experimenter after the practice session in order to stress their importance in relation to the description of the experimental procedure that was also included in the written instruction.

Each trial initially displayed three objects for two seconds. Right after that the spoken word was played while the three objects remained on the screen and a circle appeared around the critical object, the baseline object or in the empty display position for 100 msec. Each trial lasted four seconds.

The order of trials was new for each participant and pseudo-randomized with the following constraints: There were 4 sequences of three consecutive experimental items, 4 sequences of three consecutive filler items, 10 sequences of two consecutive experimental items, 10 of two consecutive filler items and 16 blocks of single experimental and filler items each. These were ordered in a random fashion, with filler and experimental item sequences alternating. Within a sequence of experimental items, there were never two consecutive items of the same condition.

Analysis

Data removal Participants gave incorrect lexical decision responses on 3.6% of all trials, which were removed from the data set. There were 110 trials with lexical decision latencies beyond two standard deviations from their mean and six additional trials which were considered to be outliers based on the inspection of scatterplots (larger than 1000 msec). All these lexical decision values were removed from the data set, resulting in missing values. Four participants did not adequately adhere to the instruction to follow the highlighting cue (they followed the cue on only 50 - 70% of trials, compared to well over 70% of trials each of the other participants) and were thus excluded from the analyses. These participants' trials were removed from the data set entirely. The resulting data set contained 86.6% of the very initial data set's trials and 82.5% of all lexical decision responses.

Two trials subsets were compiled, selecting only those trials on which participants followed the highlighting cue in order to have more homogenous data for the lexical decision analyses. One set contained trials on which the highlighted object was looked at at least once during the second before the lexical decision (88.5%) and another set contained those trials on which the highlighted object was looked at during the spoken word (88.0% of the cleaned-up data set's trials).

Inferential statistics This experiment differs from the previous experiments in that it had two factors as well as consistent item groups. Because of that it was possible to follow the advice of Jeroen Raaijmakers (personal communication, April 2009) for analyzing two-factor studies that do not have a strict Graeco-Latin Square design[34] (which is desirable for the analysis proposed in Raaijmakers et al. (1999)): First, the two factors were combined into a single one (each combination of the two original treatment factors resulting in one level of the new combined factor) and analyzed just as Raaijmakers et al. (1999) advises to analyze studies with a single factor. If these analyses showed no significant effects of both item group and combined factor*item group,

[34] In a Graeco-Latin Square design, the second factor would be nested under the first one in such a way that all the items of an item group would be seen in the same level of both factors on all subject lists.

it was justified to analyze our data with a simple ANOVA that had both factors as within-subject factors and participant list as a between-subject. If, however, either item group or combined factor*item group had significant effects on the dependent variable while participant list had not, we instead conducted an ANOVA using the two factors as within-subject factors and item group as a between-subject factor. In case that both item group or combined factor*item group and participant list significantly affected the dependent measure, there was no better option than using the standard approach of a factor 1* factor 2 * participant list ANOVA.

3.1.2 Results

The results section of this experiment is structured similarly to Experiment 1's results section. Please also refer to the 'Analysis' Section of Experiment 1 as well as to the Glossary for the meaning of specific analysis-related terms that are used in the following results section.

Deployment of visual attention

Time course of visual attention: overview Figures 3.3 and 3.4 illustrate the time course of the deployment of visual attention from word onset. At word onset, the three displayed objects received equal amounts of visual attention on all conditions, while the fourth, empty, position in the object grid was not looked at at all. The probability to look at the highlighted grid position began to rise first during the 150 - 200 msec slot after word onset on all conditions and reached peaks of above .90 at around 400 - 450 msec on those conditions on which an object was highlighted and peaks of about .65 at 250 - 300 msec when empty space was highlighted, which was well before the average word end and lexical decision responses. When the referent was present in the display but not highlighted, it drew visual attention upon perception of the spoken word: on these two conditions (the critical object is the referent and empty space or the baseline object is highlighted), looks to the referent began during the 600 - 650 msec time slot after word onset and reached inspection probabilities at the average lexical decision response of .45 when empty space was highlighted and .35 when the baseline object was highlighted.

Assessing the visual saliency of the displayed objects: Highlighted-object and critical-object inspection advantages before word onset On all conditions on which an actual object was highlighted, the highlighted-object advantages of inspection probabilities for the 'shifted time region before word onset' (cf. Figure 3.5 and Table 3.1) did not differ significantly from zero. The empty region, however, was not at all looked at before word onset: there were significant distractor advantages over highlighted-object inspection probabilities when the

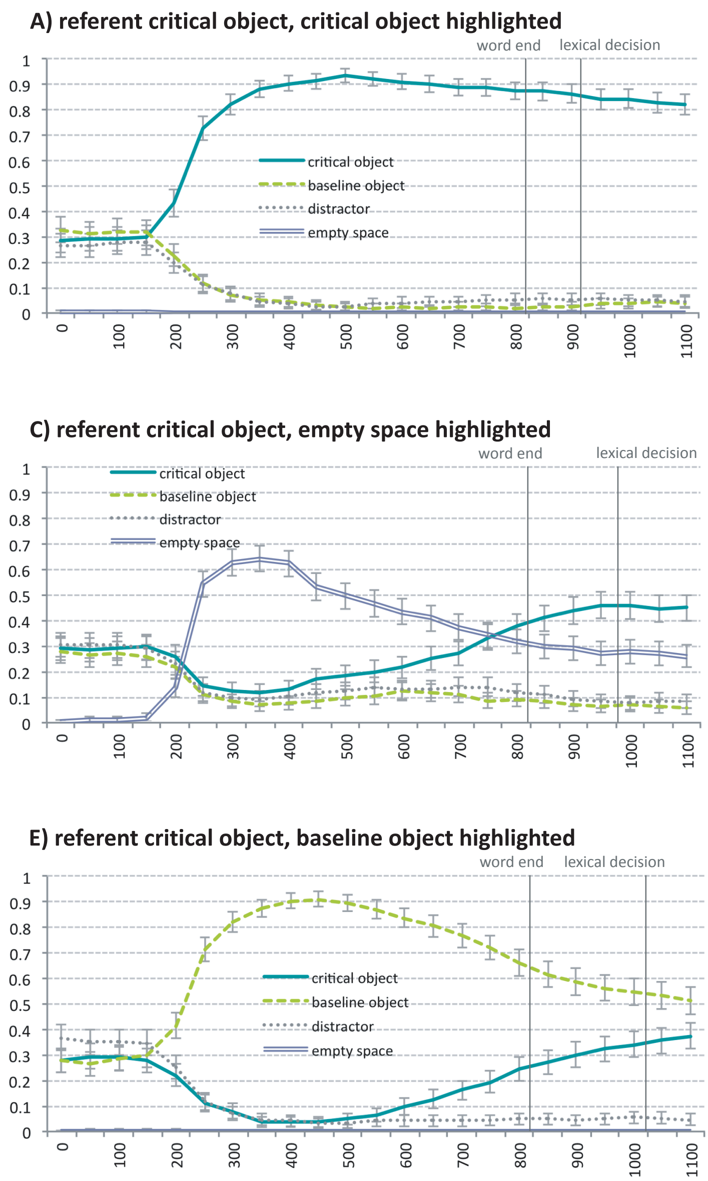

Figure 3.3: On this and the following page: Experiment 5: Critical-object, baseline-object, distractor and empty-space (i.e. the empty grid position that could be highlighted) inspection probabilities for 50-msec long time slots relative to the onset of the spoken word, with one graph (A to F) per condition. The time identifying a time slot denotes its end point. Error bars indicate 95% confidence intervals.

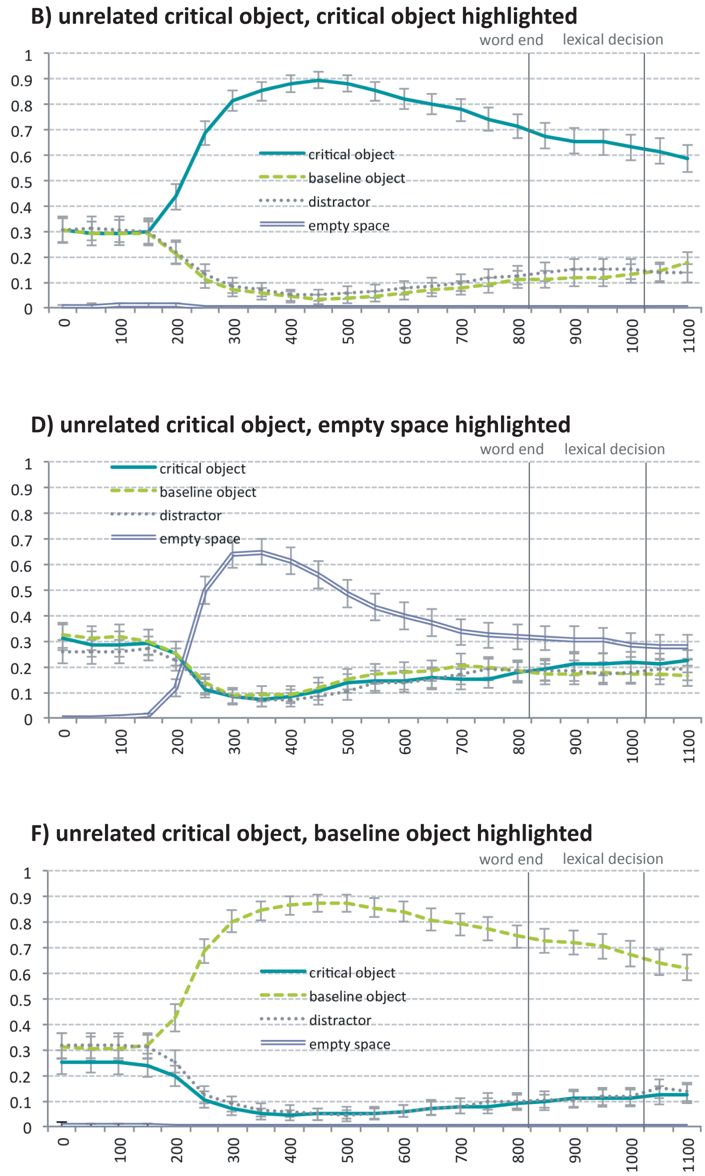

Figure 3.4: On this and the previous page: Experiment 5: Critical-object, baseline-object, distractor and empty-space (i.e. the empty grid position that could be highlighted) inspection probabilities for 50-msec long time slots relative to the onset of the spoken word, with one graph (A to F) per condition. The time identifying a time slot denotes its end point. Error bars indicate 95% confidence intervals.

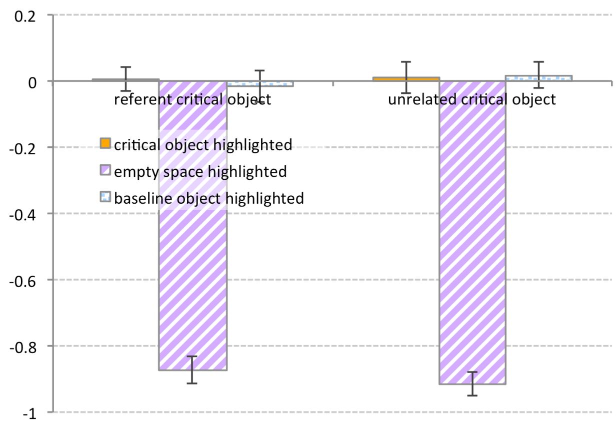

Figure 3.5: Experiment 5: Average highlighted-object advantages in probabilities of inspections having lasted before the shifted word onset, one bar per condition. Error bars indicate 95% confidence intervals.

empty region was highlighted and the critical object was the referent as well as when the empty space was highlighted and the critical object was an unrelated object.

Table 3.1: Experiment 5: Results of t-tests establishing whether highlighted-object inspection advantages before word onset were significantly different from zero.

condition	t-test by subjects	t-test by items
referent critical object and critical object highlighted	$t_1(43) = 0.33$, $p = .74$	$t_2(47) = 0.23$, $p = .819$
referent critical object and empty space highlighted	$t_1(43) = 42.9$***, $p < .001$	$t_2(47) = 47.3$***, $p < .001$
referent critical object and baseline object highlighted	$t_1(43) = .66$, $p = .51$	$t_2(47) = .69$, $p = .49$
unrelated critical object and critical object highlighted	$t_1(43) = .44$, $p = .66$	$t_2(47) = .63$, $p = .53$
unrelated critical object and empty space highlighted	$t_1(43) = 49.8$***, $p < .001$	$t_2(47) = 59.3$***, $p < .001$
unrelated critical object and baseline object highlighted	$t_1(43) = .84$, $p = .41$	$t_2(47) = .91$, $p = .37$

There were no imbalances between critical-object and distractor inspection probabilities before word onset when the critical object was not highlighted (cf. Table 3.2, the remaining two conditions in which the critical object was highlighted are covered in Table 3.1).

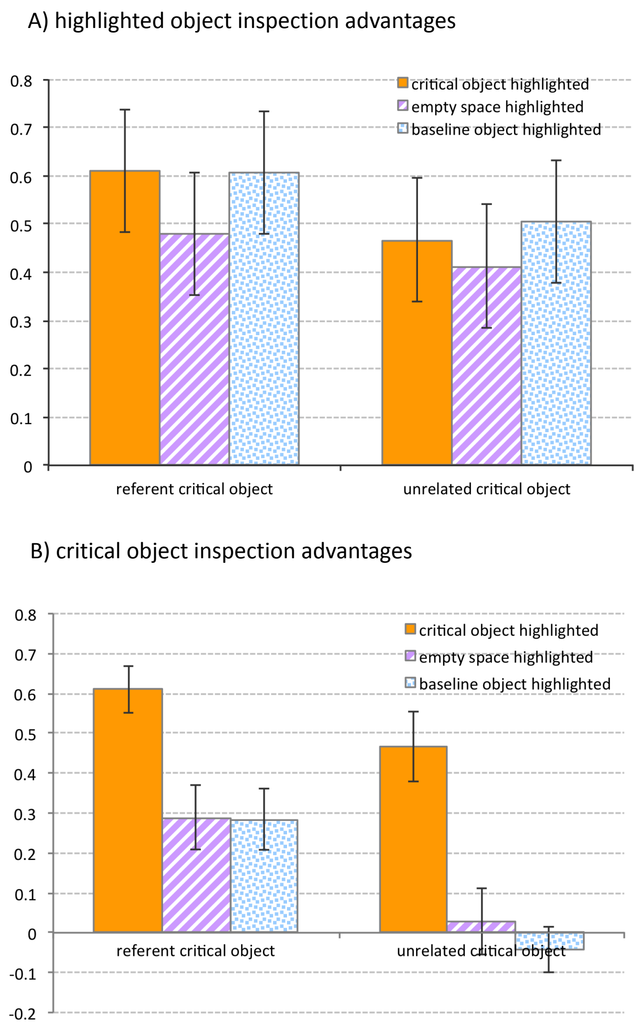

Figure 3.6: Experiment 5: Average highlighted-object (graph A)) and critical-object (graph B)) advantages in probabilities of inspections having been started during the shifted word region, one bar per condition. Error bars indicate 99% confidence intervals following Masson and Loftus (2003) for graph A) and 95% confidence intervals for graph B).

Table 3.2: Experiment 5: Results of t-tests establishing whether critical-object inspection advantages before word onset were significantly different from zero when the critical object was not highlighted.

condition	t-test by subjects	t-test by items
referent critical object and empty space highlighted	$t_1(43) = .73, p = .47$	$t_2(47) = .69\ p = .49$
referent critical object and baseline object highlighted	$t_1(43) = .12, p = .91$	$t_2(47) = .30, p = .76$
unrelated critical object and empty space highlighted	$t_1(43) = .36, p = .72$	$t_2(47) = .57, p = .57$
unrelated critical object and baseline object highlighted	$t_1(43) = .20, p = .84$	$t_2(47) = .26, p = .80$

Assessing the amount of visual attention shifts in response to the highlighting cue and spoken word onset: Highlighted-object and referent-critical-object advantages for inspections started during the spoken word As Figure 3.6 A) shows, there were significant highlighted-object advantages in probabilities of inspections having been launched during the 'shifted time region of the spoken word' on all conditions, indicating that the highlighting cue attracted visual attention reliably (cf. Table 3.3).

Table 3.3: Experiment 5: Results of one-sided t-tests establishing whether highlighted-object inspection advantages during the shifted time region of the spoken word were significantly larger than zero.

condition	t-test by subjects	t-test by items
referent critical object, critical object highlighted	$t_1(43) = 20.4^{***}, p < .001$	$t_2(47) = 19.4^{***}, p < .001$
referent critical object, empty space highlighted	$t_1(43) = 10.5^{***}, p < .001$	$t_2(47) = 13.2^{***}, p < .001$
referent critical object, baseline object highlighted	$t_1(43) = 17.6^{***}, p < .001$	$t_2(47) = 17.5^{***}, p < .001$
unrelated critical object, critical object highlighted	$t_1(43) = 10.4^{***}, p < .001$	$t_2(47) = 14.7^{***}, p < .001$
unrelated critical object, empty space highlighted	$t_1(43) = 8.4^{***}, p < .001$	$t_2(47) = 12.7^{***}, p < .001$
unrelated critical object, baseline object highlighted	$t_1(43) = 12.1^{***}, p < .001$	$t_2(47) = 15.6^{***}, p < .001$

These highlighted-object advantages were significantly affected by both the relatedness of the critical object and by the highlighted object. The main effect of the critical-object relatedness ($F(1,38) = 9.4, p < .01$) caused significantly more shifts of gaze to the highlighted referent critical object than to the highlighted unrelated critical object ($p < .05$ in pairwise comparisons). The highlighted-object main effect ($F(2,76) = 5.0$, $p < .01$) was due to less inspections being launched to the empty space than to the critical object or baseline object. Pairwise comparisons between the conditions on which the critical object was highlighted vs. on which the baseline object was highlighted showed significant differences in highlighted object advantages only when the critical object was the referent (both $p < .05$).

Table 3.4: Experiment 5: Results of t-tests establishing whether critical-object inspection advantages during the shifted time region of the spoken word were significantly different from zero on those conditions where the critical object was not highlighted.

condition	t-test by subjects	t-test by items
referent critical object, empty space highlighted	$t_1(43) = 7.1$***, $p < .001$	$t_2(47) = 7.1$***, $p < .001$
referent critical object, baseline object highlighted	$t_1(43) = 7.2$***, $p < .001$	$t_2(47) = 8.8$***, $p < .001$
unrelated critical object, empty space highlighted	$t_1(43) = .7$, $p = .50$	$t_2(47) = .9$, $p = .39$
unrelated critical object, baseline object highlighted	$t_1(43) = 1.4$, $p = .17$	$t_2(47) = 1.7$, $p = .10$

There were still significant tendencies to shift visual attention to the non-highlighted critical object during the spoken word (cf. Figure 3.6 B)) when the critical object was the referent and when the empty space or the baseline object were highlighted (cf. Table 3.4).

Assessing the potential of the highlighted object and of the critical object to affect word comprehension: Highlighted- and critical-object inspection probabilities prior to lexical decision The probabilities of highlighted-object inspections having lasted during the second before the lexical decision (cf. Figure 3.7) were significantly affected by the highlighted-object type ($F(1,38) = 69.0$, $p < .001$): The highlighted empty space was looked at less than the highlighted critical object and the highlighted baseline object with both critical object types (all $p < .001$ in pairwise comparisons). Also, the highlighted referent critical object received most looks out of all highlighted objects (all $p < .05$ in pairwise comparisons).

The highlighting factor significantly affected the proportion of inspections received by the critical object ($F(2,76) = 130.4$, $p < .001$): there were more looks to highlighted than non-highlighted critical objects (all $p < .001$ in pairwise comparisons). Also, the critical object was looked at more when the empty space was highlighted than when the baseline object was highlighted (both $p < .01$ in pairwise comparisons). The type of the critical object ($F(1,38) = 28.8$, $p < .001$) caused more looks towards the critical object when it was the referent than when it was the unrelated object ($p < .05$ when the critical object was highlighted and $p < .01$ otherwise

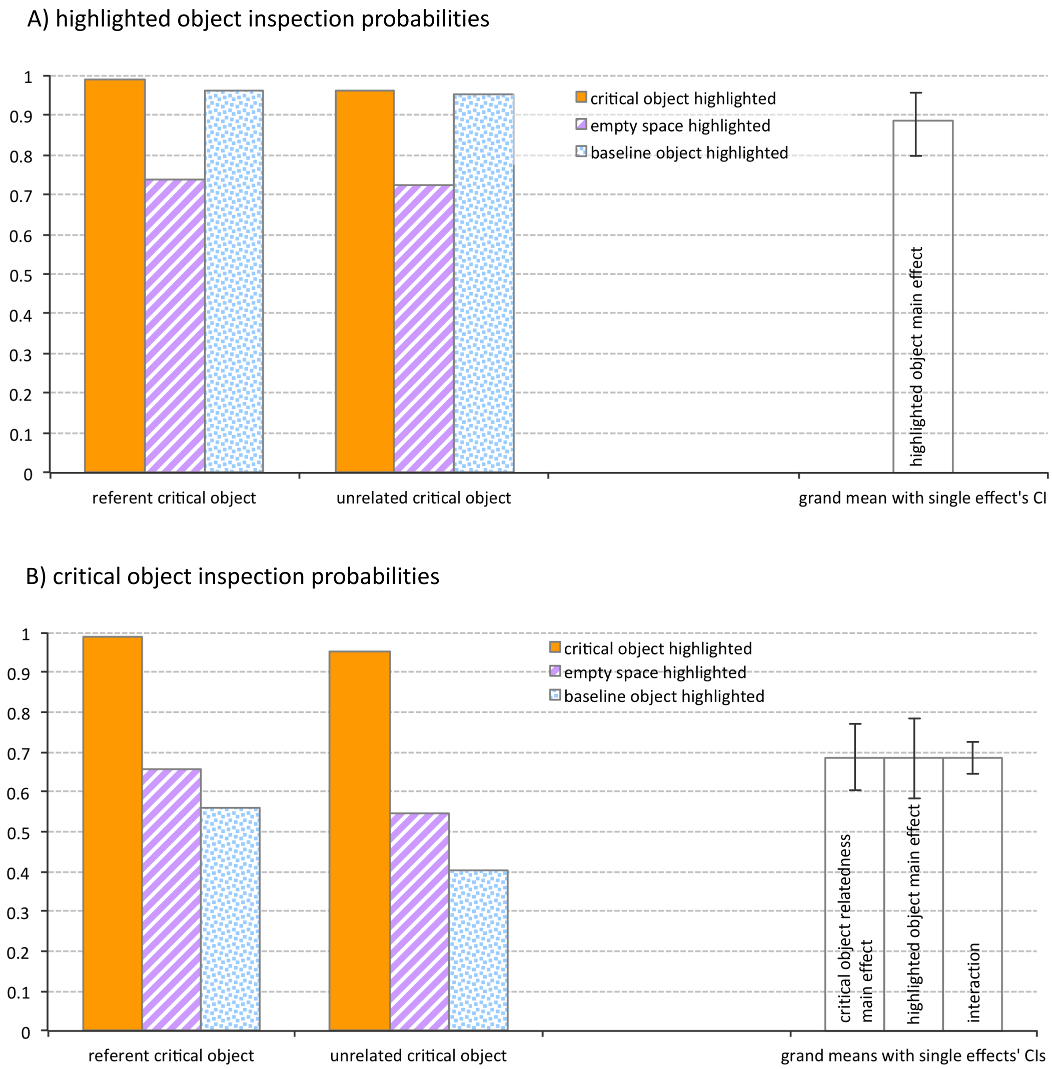

Figure 3.7: Experiment 5: Average highlighted-object (A) and critical-object (B) probabilities of inspections having lasted during the second before lexical decision, one bar per condition. For graph A), error bars indicating a 99% confidence interval for the highlighted-object main effect are plotted on a separate grand mean bar following Masson and Loftus (2003). For graph B), error bars indicating each of the main effects' (99.9%) and the interaction's (95%) confidence intervals are plotted on separate grand mean bars following Masson and Loftus (2003).

in pairwise comparisons). In addition, the interaction between critical-object relatedness and highlighted object had a significant effect ($F(2,76) = 4.0$, $p < .05$), due to the highlighted-object effect being stronger when the critical object was unrelated than it was the referent. Participants still inspected the highlighted object on over 70% of trials on all conditions, so that the preconditions for an analysis of lexical decision latencies as affected by the deployment of visual attention to the highlighted object are still fulfilled.

Speed and accuracy of lexical decision

An ANOVA on the average number of correct trials (cf. Figure 3.8) revealed a significant main effect of the factor of the critical object's type ($F(1,38) = 4.5$), $p < .05$), with more errors when the critical object was unrelated than when it was the referent. This effect was carried by significantly more correct trials when the critical object was the referent and empty space was highlighted than when the critical object was the unrelated object and empty space was highlighted ($p < .05$ in pairwise comparisons).

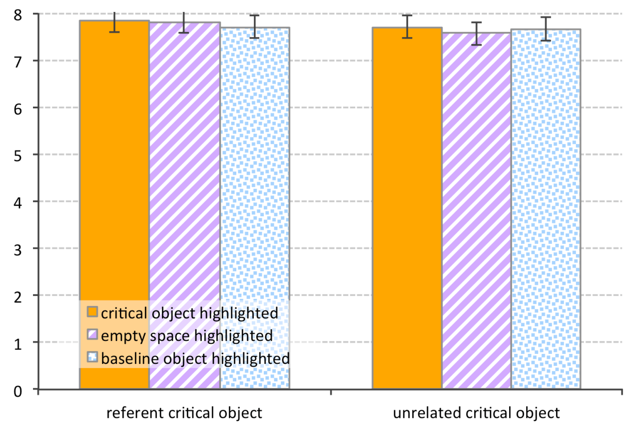

Figure 3.8: Experiment 5: Average number of correct lexical decisions per participant per condition. Error bars indicate 95% confidence intervals following Masson and Loftus (2003).

An analysis of lexical decision latencies (Figure 3.9 A)) showed several significant effects: There was a main effect of the critical-object relatedness ($F(1,38) = 53.5$, $p < .001$). This effect was carried by significant differences between the speed of lexical decision when the critical object was the referent versus when it was an unrelated object, both on the critical-object and empty-space highlighting levels (both $p < .001$ in pairwise comparisons). We also found a significant main effect of the highlighted object ($F(2,76) = 17.3$, $p < .001$). This was mainly driven by

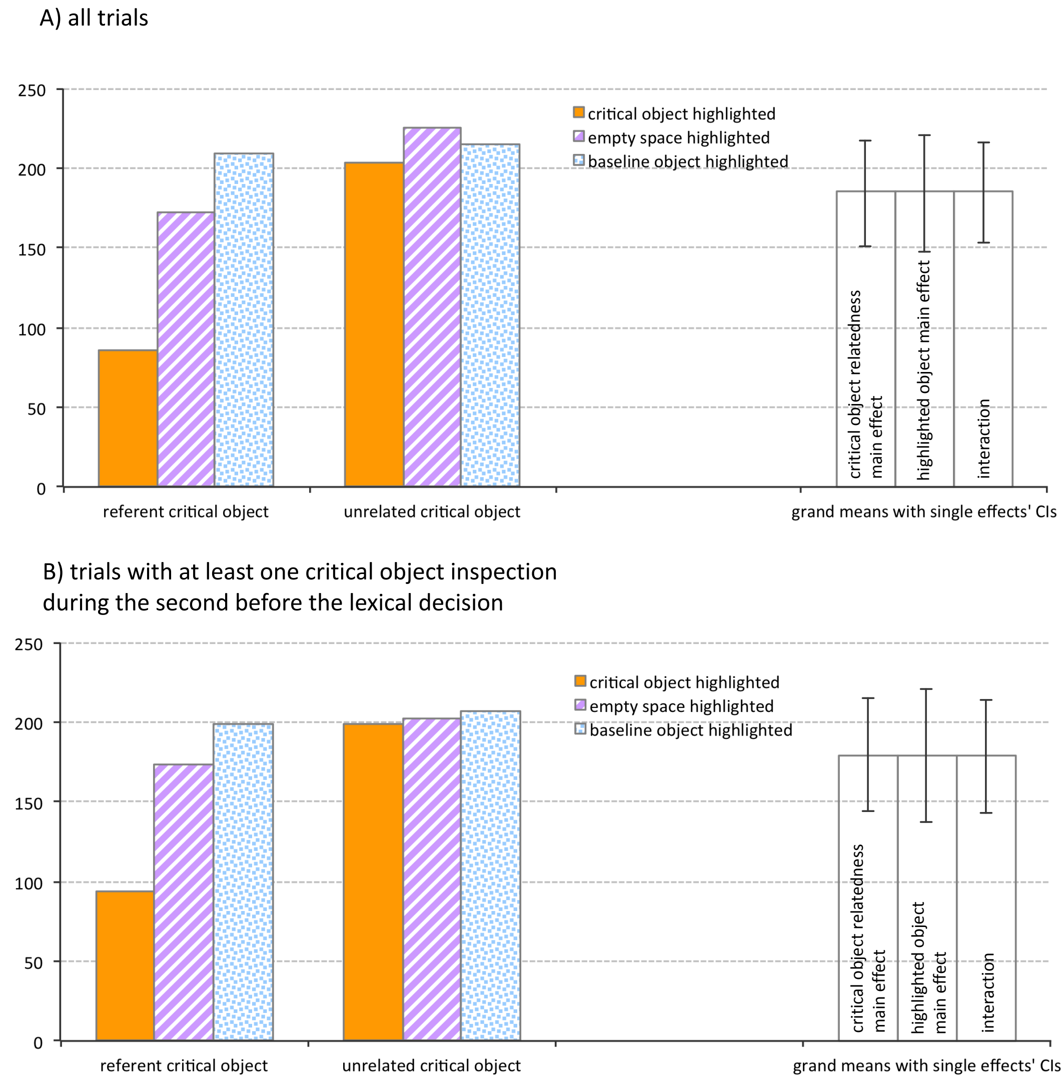

Figure 3.9: Experiment 5: Average lexical decision latencies measured from word end, one bar per condition. Error bars indicating each of the main effects' and the interaction's 99.9% confidence intervals are plotted on separate grand mean bars following Masson and Loftus (2003). Average lexical decision latencies in msec measured from word end, one bar per condition. Graphs A) and B) show averages for the entire data set and a trial subset, respectively. 99.9% confidence intervals for each effect are plotted on separate grand mean bars following Masson and Loftus (2003).

the conditions on which the critical object was the spoken word's referent. In these conditions, lexical decision latencies increase from the critical-object-highlighted (shorter lexical decision latencies when the critical object was highlighted than when empty space or the baseline object were highlighted: both $p < .001$ in pairwise comparisons) to the empty-space-highlighted and on to the baseline-object-highlighted conditions (shorter lexical decison latencies when empty space was highlighted than when the baseline object was highlighted: $p < .05$ in pairwise comparisons). The significant interaction between the two treatment factors ($F(2,76) = 14.2$, $p < .001$) meant that there was an effect of the critical object's relatedness only when the critical object was highlighted and that there was no highlighting effect when the critical object was the unrelated object. The data subset with only those trials on which participants inspected the highlighted object during the second before lexical decision (figure 3.9 B)) showed slightly different lexical decision effects: A main effect of the critical-object relatedness factor ($F(1,38) = 35.0$, $p < .001$) meant that lexical decisions were faster when the critical object was the referent than when it was unrelated. This became manifest in significantly shorter lexical decision latencies when a referent critical object was highlighted than when an unrelated critical object was highlighted (p < .001). An additional main effect of the highlighted object ($F(2,76) = 13.0$, $p < .001$) was carried by the highlighted referent critical object causing faster lexical decisions than both the highlighted empty space and the highlighted baseline object in the presence of the referent critical object (both $p < .001$). There was also a significant interaction between the two treatment factors ($F(2,76) = 12.2$, $p < .001$), which was due to the critical-object relatedness main effect affecting lexical decision only when the critical object was highlighted and the highlighted object main effect acting only on lexical decision latencies when the critical object was a referent. Analyses of the data subset containing only those trials on which the highlighted object was looked at during the spoken word yielded exactly the same pattern in lexical decision latencies, thus we don't report these in detail.

3.1.3 Discussion

With respect to the deployment of visual attention in the display, we identify three key findings: Firstly and most importantly, participants followed the visual highlighting cue reliably on all conditions - they looked at the highlighted region quickly after word onset on most trials. Secondly, the empty space was only looked at when and as soon as it was highlighted, and even then it received fewer looks than highlighted *objects*. Thirdly, the referent critical object received visual attention after word onset even when it was not highlighted. For the 'referent critical object and empty space highlighted' condition this is not surprising, as empty space is not an attractive target for visual attention given the co-presence of actual objects in the display. The referent inspection advantage on the 'referent critical object and baseline object highlighted' condition could be explained by the short duration of the highlighting cue: from 100 msec after word onset, it is no longer present around the baseline object to keep the referent from pulling attention away, which participants would prefer to look at when hearing the referent's name.

Regarding lexical decision latencies, most importantly, we found a stable and significant speed-up of lexical decisions when participants looked at the spoken word's referent during the comprehension of the spoken word, compared to all other conditions.

The entire data set also showed a lexical decision benefit on the condition on which the empty space was highlighted and the critical object was the spoken word's referent. This benefit was not significant for the subset with only those trials on which participants inspected the highlighted object during the second before the lexical decision response though, which suggests that this lexical decision speed-up in the entire data set can be attributed to participants' significant tendency to shift their gaze towards the referent after having followed the highlighting cue on this condition. Therefore, we will attribute this effect to our incomplete control over participants' deployment of visual attention and base our conclusions on lexical decision latencies being faster when the referent critical object was highlighted than on any other condition, and equally fast on all those other conditions. We conducted an additional experiment that supports this view (not reported in full here because it had only just been finished at the time of writing). This experiment was very similar to Experiment 5 and had only three conditions: The display always contained the spoken word's referent and either this referent, empty space or the unrelated baseline object were highlighted from word onset. Crucially, participants' visual attention was controlled continuously by a 'spotlight' that was on the display at all times, highlighting the grid position that participants were supposed to look at at a given point in time. This spotlight highlighted all three displayed objects as well as the empty grid position one after another before word onset, and then highlighted one grid position continuously from word onset until the end of the trial. Participants followed this cue very reliably and thus referent looks when the referent was not highlighted, which could confound the lexical decision process, became less likely. Lexical decision latencies in this experiment, where visual attention was controlled strictly, confirm our interpretation of Experiment 5's results: Lexical decision latencies were shorter when the referent was looked at during the spoken word, but equally long when empty space or the unrelated baseline object were highlighted.

In order to determine whether the 'preservation of priming' hypothesis is supported by our results, the 'referent critical object and empty space highlighted' condition plays the most crucial role. The above discussion argues for the lexical decision latencies on that condition being similar to those on the 'referent critical object and baseline object highlighted' condition, which means that this hypothesis was not confirmed by our results.

With equally fast lexical decision latencies on both empty-space-highlighted conditions as on the baseline-object-highlighted conditions we can also state that we have found no evidence for the 'avoidance of interference' hypothesis. Crucially, this relies not only on those conditions on which the referent object was present in the display and attracted some visual attention despite our highlighting cue. An interference from seeing an unrelated object during noun comprehension should also manifest itself in a difference between the conditions on which the referent object was never in the display and empty space vs. an unrelated object was looked at from word onset. As this was not the case, our results clearly lack evidence supporting the avoidance-interference hypothesis.

Lastly, since lexical decision responses were not given more quickly on all those conditions on which the referent was present than on those conditions on which the critical object was unrelated to the spoken word, we can state that our experiment (in unison with Experiment 3) does not support the idea that the mere encounter of a spoken word's referent before that word's comprehension is sufficient to facilitate word recognition. Rather, our results support the notion that listeners must fixate a spoken word's referent during the word's comprehension in order for the referent's concept to have a beneficial effect on word recognition.

3.2 Summary

The first four experiments of this thesis showed evidence that looks towards related objects during the comprehension of a spoken noun affect the speed of that noun's comprehension. Specifically, it is beneficial to look at a referent object rather than at an unrelated object while perceiving the spoken noun. In a first stab at identifying exactly which mechanisms lie behind this benefit, Experiment 5 manipulates whether a spoken noun's referent is displayed at all as well as whether participants' visual attention is directed towards that referent object (if it is shown), towards an unrelated object or towards an empty region of the display. When the referent was shown in the display, participants gave faster lexical decision responses to the spoken noun when a visual cue attracted their gaze to the referent at word onset than when the cue highlighted empty space or an unrelated object, replicating Experiment 3's results. Also, on all those conditions on which the referent was not displayed, lexical decisions were as slow as when the referent was displayed but an unrelated object was highlighted. As laid out in the above discussion, these results provide no evidence for either the hypotheses that NCEs could benefit noun comprehension by preserving a priming effect that originates in having seen the noun's referent before noun onset or that looks to a referent would be beneficial by avoiding interference that would arise from looking at unrelated objects during noun comprehension. Instead, Experiment 5 shows that it is the simultaneous perception of the visual referent image and the spoken noun that makes noun-contingent referent looks beneficial for noun comprehension. The following chapter will use another theoretical framework entirely in order to explore potential mechanisms behind a benefit of noun-contingent referent looks on noun comprehension.

4 Access to referent location as the source of a referent-look comprehension benefit

Up to this point, this thesis has shown empirically that seeing a depicted object while hearing its name speeds up the recognition of the object name. Experiment 5 specifically has shown that it is the simultaneous perception of an object's visual image that is beneficial for the comprehension of its name. It is clear that an eye movement to the location of a previously inspected object in a display will, in most situations, provide one with the actual visual image of that object. Beyond this, however, the spatial indexing framework supports the idea that accessing the *location* of a previously seen referent through an eye movement could provide information that is beneficial for the comprehension of a spoken noun as well. In Section 1.5 we propose, following a spatial indexing account, that the objects seen during the preview period of a visual-world study are encoded with their locations and that looks back to these location activate the associated internal representations of the respective objects. Noun-contingent referent looks could thus be beneficial to noun comprehension by providing access to the internal representation of the noun's referent, as created during the preview period and accessed through a look to the referent's *location*. Experiments 6 and 7 search for empirical evidence for the existence of such a mechanism.

4.1 Bare location access: Experiment 6

Chapter 3 showed that the comprehension of a spoken noun is facilitated by looks to the noun's referent during comprehension. As described in Section 1.5, it might be hypothesized that not only access to the spoken noun's visual representation, but also access to a location in the environment in which this visual representation has been perceived earlier could be beneficial for noun comprehension. Experiment 6 tests this hypothesis. After having had the opportunity to inspect three objects on the display, participants heard a spoken word that referred to one of these objects. At word onset, the display went blank and only a circle remained on screen for a very short period of time. This circle highlighted either the referent's initial position or the position of an object unrelated to the spoken word (cf. Figure 4.1).

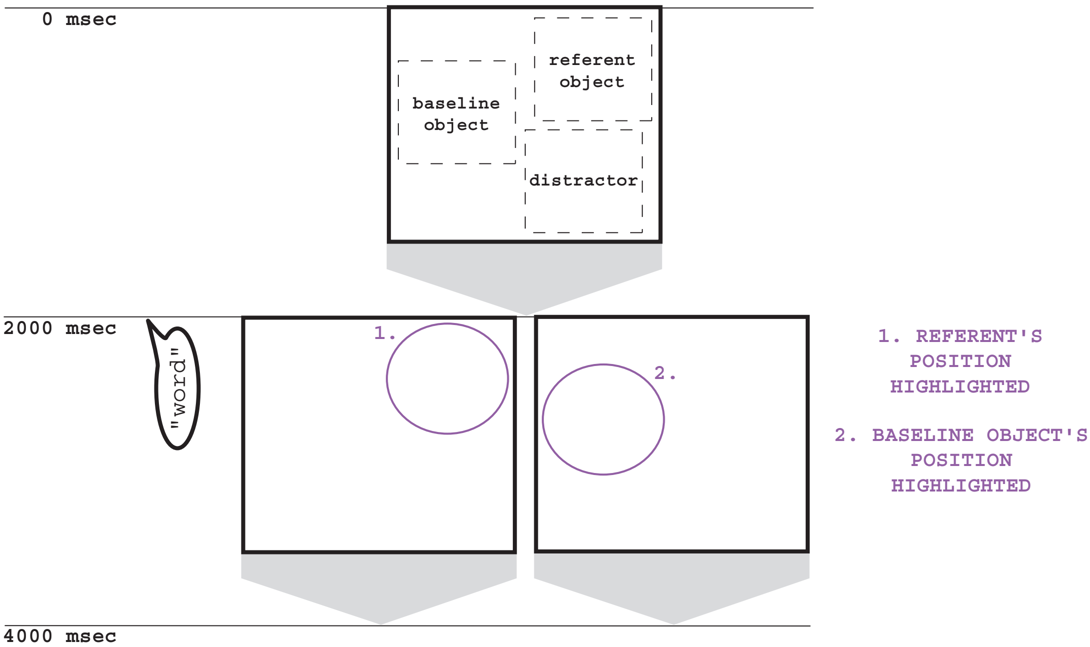

Figure 4.1: A sample trial of Experiment 6. The display, containing the spoken word's referent, an unrelated baseline object and one distractor, was shown with a preview of two seconds. Then, the spoken word began to play while the display changed. The new display contained only a circle, which could surround either the referent's or the baseline object's initial position. Participants gave their lexical decision response via a button press. Every trial had an overall duration of four seconds.

Following the encoding of objects and their locations during the preview period, the strong visual cue of the single circle on the display should direct the eyes to the initial location of either the referent or the baseline object during the lifetime of the spoken noun. These eye movements could potentially provide information that could interact with comprehension. A benefit from accessing the referent's initial location would show in faster lexical decision latencies on the condition on which the referent's initial location was highlighted compared to when the baseline object's initial location was highlighted.

4.1.1 Method

Participants

Twenty native German speakers from Saarland University with normal or corrected-to-normal vision took part in this experiment for a payment of five Euros. The data of four of them was not included in the analyses because they either guessed the experiment's purpose or hardly ever followed the highlighting cue, as assessed by the experimenter while running the experiment. The resulting data set used for the analyses comprised the data of 16 participants.

Materials

This experiment's items were drawn from the 24 experimental items used in Experiment 3. The types of objects in the two experiments are related as follows: In Experiment 6, an item's spoken word and referent remained the same as in Experiment 3, the unrelated critical object from Experiment 3 became the unrelated baseline object of this experiment, and Experiment 3's baseline object was used as a distractor in this experiment.

Each display initially contained three objects: the spoken word's referent, the baseline object and a distractor (cf. Figure 4.2). These objects were arranged in a triangle, using the same eight triangular grid templates as in Experiment 3. This ensured that, again, the critical objects (referent as well as baseline object) appeared equally often in each screen quadrant across items.

The spoken stimuli were the same as in Experiment 3 and had an average duration of 622 msec across experimental trials.

In addition to the 24 experimental items there were 48 filler items, which resembled the experimental items in their trial procedure. They consisted of 12 word and 36 non-word items, out of which six had phonological overlap between one displayed object and the spoken word. This object's location was highlighted on some trials. Before the actual start of the experiment, there were ten practice trials similar to the experiment's item set.

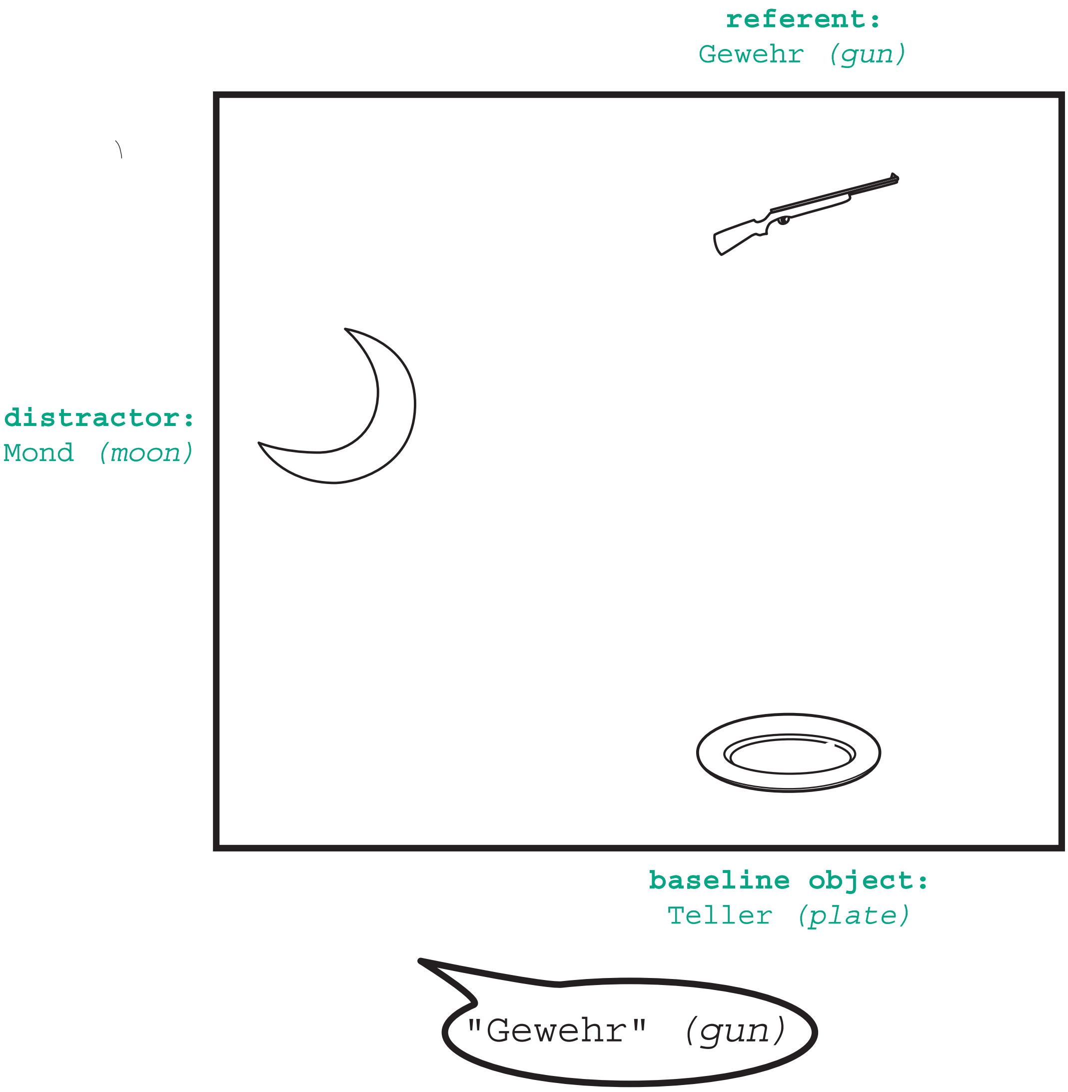

Figure 4.2: The materials for one of Experiment 6's items. Participants were presented with a display containing three inanimate objects: the spoken noun's referent, an unrelated baseline object and a distractor. The display remained as shown here from trial onset until the onset of the spoken noun. At that point, the display went blank but for a circle around one object's previous position (cf. Figure 4.1).

Design

This experiment had a single-factor (circle location, which had two levels) within-subject design.

Procedure

The equipment used in this experiment was the same as in Experiment 5. In the instruction text, participants were asked to inspect the three displayed objects before word onset and to then make their lexical decision on the spoken stimulus as quickly and accurately as possible by pressing a button. Just before the start of the experiment itself, they were also reminded by the experimenter to attend to the highlighting circle. Each trial began with a display (containing the spoken word's referent, the baseline object and a distractor) preview that lasted two seconds, right after which the word was played while the screen switched to display only the circle. The overall trial duration was four seconds.

The order of trials was different for each participant and random except for the constraint that each experimental item was preceded by two filler items.

Analysis

Data removal Fourteen trials were removed from the initial data set due to inaccurate lexical decision, leaving 96% of trials. Seventeen trials lay beyond two standard deviations from their means and an additional five lexical decision latencies lay outside subjectively (through scatter-plot inspection) determined boundaries of -200 and 600 msec. These latencies, making up 5.8% of all trials, were removed from the data set.

The data subsets containing only those trials on which participants inspected the highlighted position during the word, while they were making the lexical decision, or during the second before the lexical decision were all large (containing above 85% of all trials), because participants' attention was caught by the highlighting cue reliably. Due to the subsets' similar size to the overall pre-processed data set described above, they were not used for separate lexical decision analyses.

Inferential statistics As this experiment had a single factor with two levels, the advised analysis is an ANOVA with the treatment as a within-subject factor and participant list as a between-subject factor (Raaijmakers et al., 1999).

4.1.2 Results

The results section of this experiment is structured similarly to Experiment 1's results section. Please also refer to the 'Analysis' Section of Experiment 1 as well as to the Glossary for the meaning of specific analysis-related terms that are used in the following results section.

Deployment of visual attention

Time course of visual attention: overview Participants' viewing behavior was very similar on both conditions (cf. Figure 4.3): Before word onset, all three objects in the display were looked at at chance level on both conditions. Looks to the highlighted position were started first during the 200 - 250 msec slot after word onset, and the probability to look at the highlighted position reaches its peak (inspection probability of .9) at around word end.

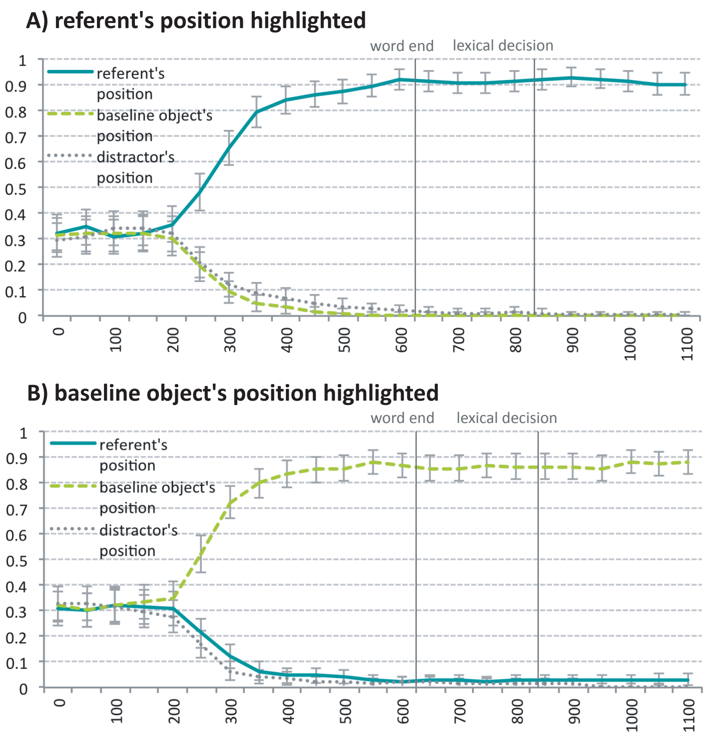

Figure 4.3: Experiment 6: Referent's, baseline object's and distractor's position inspection probabilities for 50-msec long time slots relative to the onset of the spoken word, with one graph (A and B) per condition. The time identifying a time slot denotes its end point. Error bars indicate 95% confidence intervals.

Assessing the visual saliency of the displayed objects: Highlighted-position and referent inspection advantages before word onset T-tests confirmed that before the shifted word onset, the probabilities to inspect the highlighted object and referent did not differ significantly from the distractor inspection probabilities on either condition (highlighted-position advantage when referent position would be highlighted: $t_1(15) = .6$, $p = .58$, $t_2(23) = .8$, $p = .45$; highlighted-position advantage when baseline object's position would be highlighted: $t_1(15) = .3$, $p = .78$, $t_2(23) = .3$, $p = .75$; referent advantage when baseline object's position would be highlighted: $t_1(15) = .3$, $p = .75$, $t_2(23) = .3$, $p = .77$).

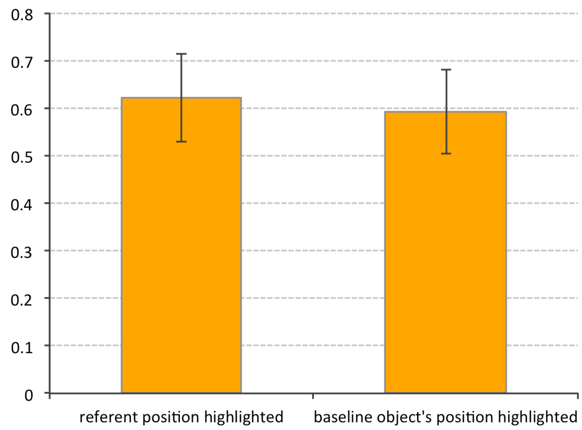

Figure 4.4: Experiment 6: Average highlighted-position advantages in probabilities of inspections having been started during the shifted word region, one bar per condition. Error bars indicate 95% confidence intervals.

Assessing the amount of visual attention shifts in response to the display change and spoken word onset: Highlighted-position and referent-position advantages for inspections started during the spoken word For the time region of the shifted spoken word, there were significant highlighted-position inspection advantages (cf. Figure 4.4) on both conditions (referent position highlighted: $t_1(15) = 13.1$, $t_2(23) = 12.8$; both $p < .001$; baseline object's position highlighted: $t_1(15) = 13.1$, $t_2(23) = 11.2$; both $p < .001$), with no significant effect of the highlighting factor ($F(1,12) = .3$, $p = .58$). There was no significant tendency to shift visual attention to the referent's position during the spoken word when the baseline object's position was highlighted ($t_1(15) = .3$, $p = .75$, $t_2(23) = .2$, $p = .88$).

Assessing the potential of the highlighted position and referent position to affect word comprehension: Highlighted-position and referent-position inspection probabilities prior to lexical decision The probability to look at the highlighted position during the second before the lexical decision response was on average .97 and was not affected by the highlighting factor ($F(1,12) = .5$, $p = .5$). The probability to look at the referent's position during this time region was significantly affected by the highlighted position factor ($F(1,12) = 304.0$, $p < .001$), with more looks to this position when it was highlighted than when it was not. When the baseline object's position was highlighted, there were still inspections to the referent's position during the second before the lexical decision response on 42% of trials.This seems to stand in contrast to Figure 4.3, which shows almost no post-word onset looks to the referent position when it was not highlighted. The time region of the second before the lexical decision, however, sometimes started before word onset (when the lexical decision happens less than one second after word onset) and thus included some of the much more frequent referent looks made during the preview period. Importantly though, the amount of referent looks during the second before the lexical decision was almost exactly as high as the amount of distractor looks during this time region (also cf. the preceding paragraph).

Speed and accuracy of lexical decision

The accuracy of lexical decisions was on average 98.3% (5.9 out of six trials per participant and condition) and was not affected by the highlighting factor ($F(1,12) = 1.5$, $p = .25$). Lexical decision latencies (Figure 4.5) were not affected by the highlighting factor ($F(1,12) = 304.0$, $p < .001$).

4.1.3 Discussion

In sum, participants looked at the displayed circle during the spoken word and before the lexical decision response very reliably and equally so on both conditions. The spoken word did not cause shifts of gaze towards the referent's empty position when it was not highlighted.

This experiment offers no support for the notion that the mere re-enactment of a previous referent look or access to the location of a previously seen referent would benefit the comprehension of a spoken noun. Although it is still conceivable that the circle displayed from word onset was taken as mismatching visual information that interfered with a potential location benefit, this is improbable as the circle was the same on every trial (making it less likely to be regarded as an object that could be related in some way to the spoken word instead of as a marker of where to look) and surrounded rather than covered up the area of the original objects.

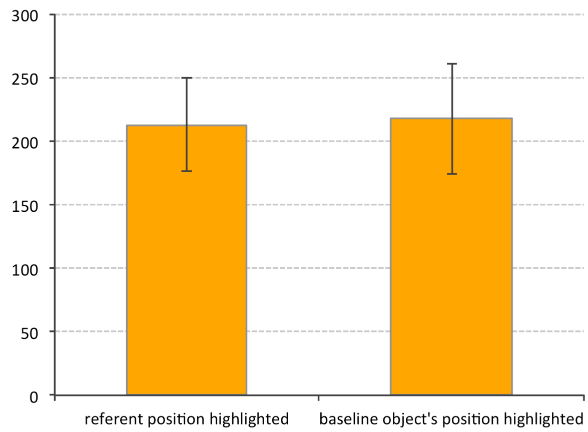

Figure 4.5: Experiment 6: Average lexical decision latencies in msec measured from word end (entire data set), one bar per condition. Error bars indicate 95% confidence intervals.

4.2 Image contents vs. location access: Experiment 7

The first experiment described in this chapter produced results in line with those from Hollywood-Squares studies (e.g. Richardson and Spivey (2000)): As in recall, noun comprehension does not benefit from looks to the *empty* locations of relevant objects (cf. Section 1.5 for a detailed comparison of our question to the specific setting of Hollywood Squares studies). Experiment 7 continues to search for evidence of spatial indexing as part of the comprehension benefit of noun-related eye movements. It asks whether an eye-movement to the location of a previously-seen referent could be beneficial for noun comprehension when that location is not empty, but when it contains an image. The perception of this image is likely to have its own effect on the comprehension of a spoken noun, which may be stronger than the effect of location access. In order to tease apart these two factors, Experiment 7 directed participants' visual attention towards different locations (where the referent had been seen during the preview period vs. where an unrelated baseline object had been seen) and independently provided different images at these locations (the noun's referent vs. the unrelated baseline object).

The location to which the eyes were drawn during the spoken word's comprehension as well as the visual image provided at this location were manipulated. The display initially contained three objects: the spoken word's referent, an unrelated (lexical decision) baseline object and a distractor. After a preview period the spoken word was played for lexical decision. At word

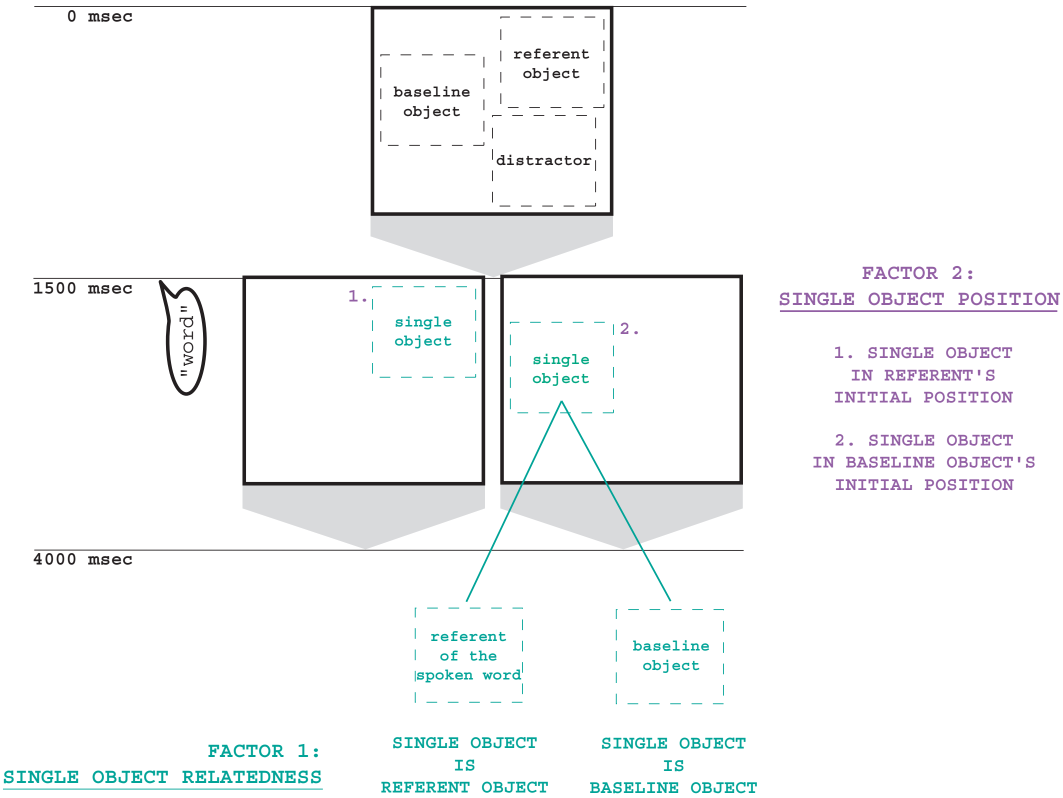

Figure 4.6: A sample trial of Experiment 7. The display, containing the spoken word's referent, an unrelated baseline object and one distractor, was shown with a preview of about 1500 msec. Then, the spoken word began to play while the display changed. The new display contained only a single object (the referent / the baseline object), which could be located either in the referent's or the baseline object's initial position. Participants gave their lexical decision response via a button press. Every trial had an overall duration of four seconds.

onset, the display changed to contain only a single object, which was either the spoken word's referent or the unrelated baseline object. This single object was displayed in either the referent's or the baseline object's initial display position (cf. Figure 4.6).

We expected participants to follow our instruction, which was to inspect all three displayed objects prior to word onset and to then shift gaze to the display location of the single object while the spoken word was played. This way the location participants directed their visual attention to (the position they initially saw the referent in vs. the position they initially saw an unrelated object in) as well as the visual image they perceived at this location (the referent vs. an unrelated object) were manipulated. It was thus possible to analyze the effects of two kinds of information provided by noun-contingent referent looks on noun comprehension: the location (and information associated to it) and the visual entity brought into focus by looking at them.

4.2.1 Method

Participants

Thirty native German speakers from the Saarland University community participated in this experiment. They had normal or corrected-to-normal vision and were paid five Euros. Three participants' data was not included in the analyses because they rarely looked at the single object or were very unfocused throughout the experiment. Overall, 27 participants' data contributed to the analyses.

Materials

The materials, visual as well as linguistic, were identical to those in Experiment 6 (cf. Figure 4.7, which is included here for your convenience, although its contents are identical to Figure 4.2).

Design

This experiment had a 2 (single-object relatedness) x 2 (single-object position) within-subject design, with four consistent participant lists and four item groups.

Procedure

Experiments 6 and 7's procedures were largely similar, the only difference between them being the stimulus presentation: In Experiment 7, each trial began with a preview of the three initially displayed objects that lasted about 1500 msec. Immediately after the preview, the spoken word began at the same moment at which the display was replaced by a new display containing only

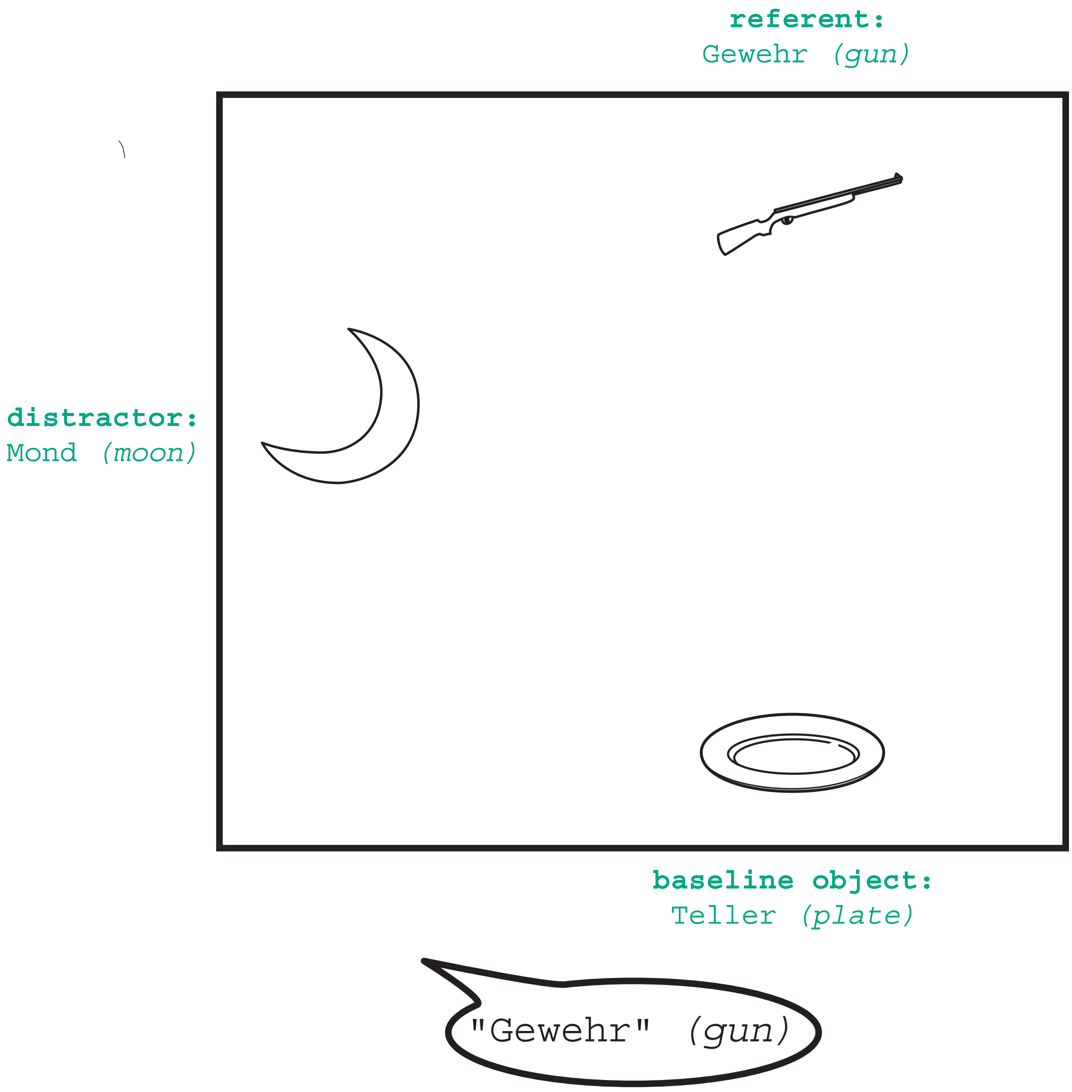

Figure 4.7: The materials for one of Experiment 7's items, which are exactly the same as for Experiment 6. Participants are presented with a display containing three inanimate objects: the spoken noun's referent, an unrelated baseline object and a distractor. The display remains as shown here from trial onset until the onset of the spoken noun. At that point, the display went blank but for a single object (cf. Figure 3.1).

a single object - its location and identity depending on the condition (cf. Figure 4.6). Each trial lasted four seconds.

Analysis

Data In this experiment, there were two objects or locations on the screen which could attract visual attention from the onset of the spoken word: The single object as well as the referent's initial position (which could be empty from word onset but had previously contained the spoken word's referent). The distractor object constituted the visual baseline and was thus used to compute single object inspection advantage measures against. It is important to note, in the following paragraphs, the distinction between the notions of the single object and of the referent or baseline object's initial positions: The two latter terms are straightforward and refer to the constant display locations in which the referent and baseline objects could be perceived in during the preview period. In contrast, the term 'single object' refers to the display location in which the single object is displayed after the spoken word's onset, which will be the referent's or the baseline object's initial position, depending on the condition.

Data removal 1.4% of the initial data set's trials were removed due to incorrect or premature lexical decision responses. In the resulting data set, 13 lexical decision latencies lay beyond two standard deviations away from their means, and seven additional lexical decision latencies were judged to be outliers based on scatterplot inspection (being lower than -200 or higher than 750 msec). These latencies (on 3.2% of the initial data set's trials) were removed from the data set. No further data needed to be removed on the basis of participants' viewing behavior, as all participants looked at the single object reliably. Thus, the data set used in the visual attention and overall lexical decision latency analyses contained 98.6% of all trials, and 95.4% of all lexical decision responses. The single object was looked at during word comprehension on so many trials (on above 94% of all trials) that there was no need to analyze lexical decision latencies for these trial subsets.

Inferential statistics As this experiment had two within-subject factors and consistent item groups, inferential statistics were conducted as in Experiment 5, which had a similar design.

4.2.2 Results

The results section of this experiment is structured similarly to Experiment 1's results section. Please also refer to the 'Analysis' Section of Experiment 1 as well as to the Glossary for the

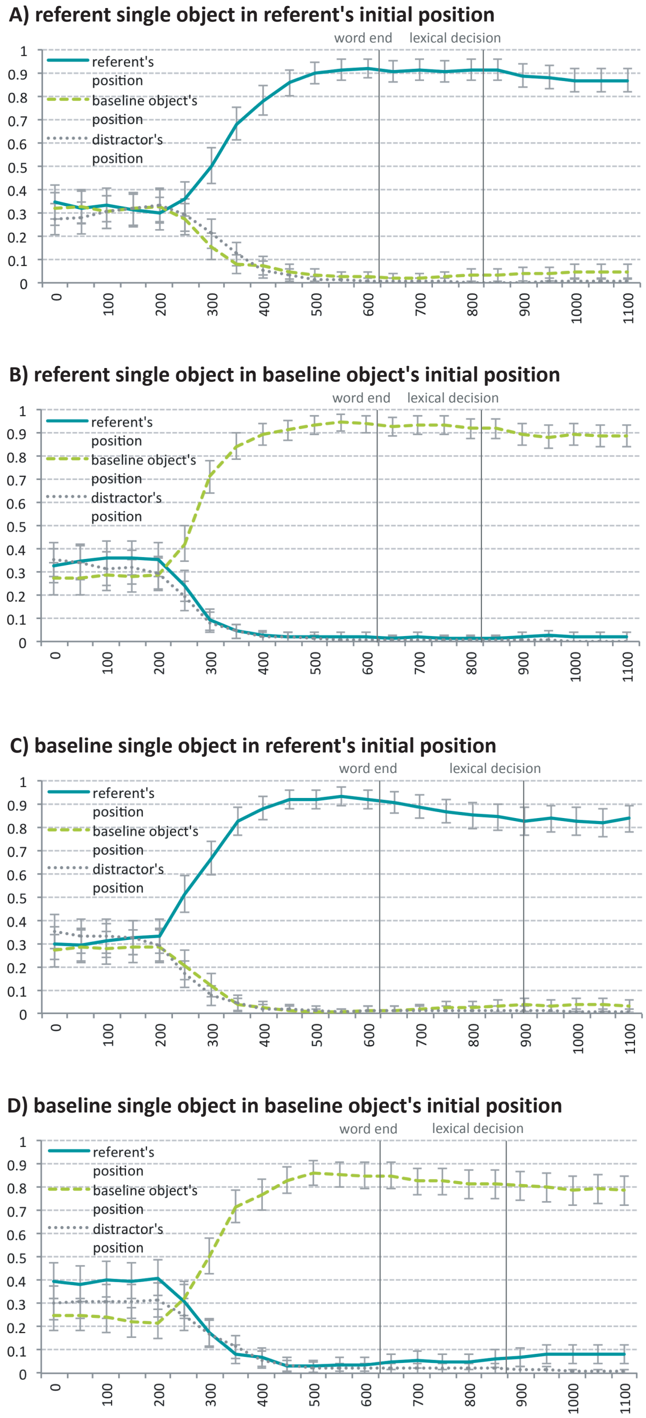

Figure 4.8: Experiment 7: Average referent-position, baseline-object-position and distractor-position inspection probabilities for 50-msec long time slots relative to the onset of the spoken word, with one graph (A to D) per condition. The time identifying a time slot denotes its end point. Error bars indicate 95% confidence intervals.

meaning of specific analysis-related terms that are used in the following results section.

Time course of visual attention: overview Figure 4.8 shows that there were, at word onset, about equally high inspection probabilities for the referent-object, baseline-object and distractor positions on all conditions apart from the 'baseline object in its own initial position' condition. Looks to the single object were started first during the 200 - 250 msec slot after word onset on all conditions and peaked at a single-object inspection probability of above .9 during the 500 - 550 msec time slot (when the single object was the referent or when it was the baseline object in the referent's initial position) or of about .85 during the 450 - 500 msec time slot (when the single object was the baseline object in its own initial position). This means that single object inspection probabilities reached their peak between 50 and 150 msec before the end of the spoken word. Also, the referent's position attracted looks even when there was no referent present during the spoken word: when the single object displayed from word onset was the baseline object in its own initial position, the probability to look at the empty referent's position remained about .1 after word onset.

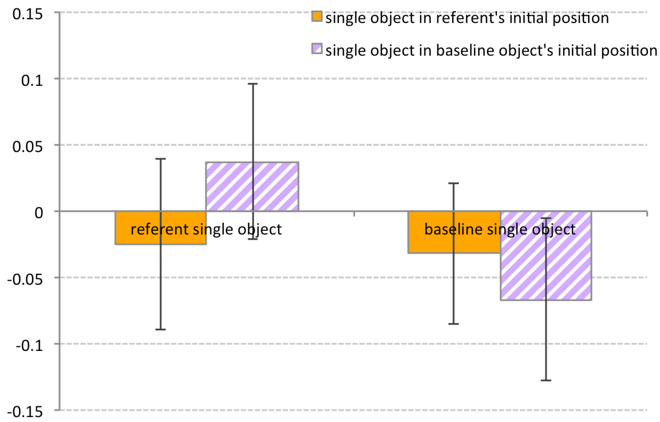

Figure 4.9: Experiment 7: Average single-object advantages in probabilities of inspections having lasted before the shifted word onset, one bar per condition. Error bars show 95% confidence intervals.

Assessing the visual saliency of the displayed objects: Single-object and referent inspection advantages before word onset The slight tendency to look at the distractor more than at the baseline object during the preview period of the 'baseline object in its own initial position' condition, hinted at in the time course of visual attention, was confirmed by a one-sample t-test on single-object (which is in the baseline object's initial position in this condition) inspection advantages for the time region before the shifted word onset (cf. Figure 4.9): There was a marginally significant distractor advantage over single-object inspection probabilities when the single object was the baseline object in its own initial position ($t_1(26) = 2.1$, $p < .05$; $t_2(23) = 1.8$, $p = .09$). There were no further single-object, referent or distractor advantages before word onset on other conditions (referent in its own initial position, single object vs. distractor inspections: $t_1(26) = .75$, $p = .46$; $t_2(23) = .66$, $p = .52$; referent in the baseline object's initial position, single object vs. distractor inspections: $t_1(26) = 1.2$, $p = .23$; $t_2(23) = .99$, $p = .33$; referent in the baseline object's initial position, referent vs. distractor inspections: $t_1(26) = 0$, $p = 1$; $t_2(23) = .27$, $p = .79$; baseline object in referent's initial position, single object vs. distractor inspections: $t_1(26) = 1.2$, $p = .25$; $t_2(23) = .80$, $p = .43$; baseline object in its own initial position, referent vs. distractor inspections: $t_1(26) = .13$, $p = .90$; $t_2(23) = .57$, $p = .58$).

Assessing the amount of visual attention shifts in response to the display change and spoken word onset: Single-object and initial-referent-position advantages for inspections started during the spoken word The single object clearly attracted shifts of attention on all conditions (Figure 4.10): One-sample t-tests on single object advantages in the probabilities of inspections started during the shifted word region show significantly positive single object advantages on all conditions ($t_1(26)$ between 14.1 and 21.3; $t_2(23)$ between 16.5 and 22.1; all $p < .001$ (one-tailed)). The single object's position significantly affected these single-object advantages ($F(1,23) = 4.3$, $p < .05$), with the single object in the baseline object's initial position receiving more looks than the single object in the referent's initial position numerically, but pairwise comparisons not showing significant differences between specific condition pairs (all $p > .11$). The referent's initial position was not looked at at all when the single object was the referent in the baseline object's initial position, but received a small, marginally significant amount of looks when it was the baseline object in its own initial position ($t_1(26) = 1.5$, $p = .05$ (one-tailed); $t_2(23) = 1.8$, $p = .05$ (one-tailed)).

Assessing the potential of the single object and referent's initial position to affect word comprehension: Single-object and referent-position inspection probabilities prior to lexical decision The probability for participants to have inspected the single object at least once during the second before the lexical decision response was on average .97 and was not affected by the treatment factors (single-object relatedness: $F(1,23) = 0.3$, $p = .60$; single-object position: $F(1,23 = .004$, $p = .05)$; interaction: $F(1,23) = .4$, $p = .56$). The probability to have inspected the referent object's position at least once during that time period was

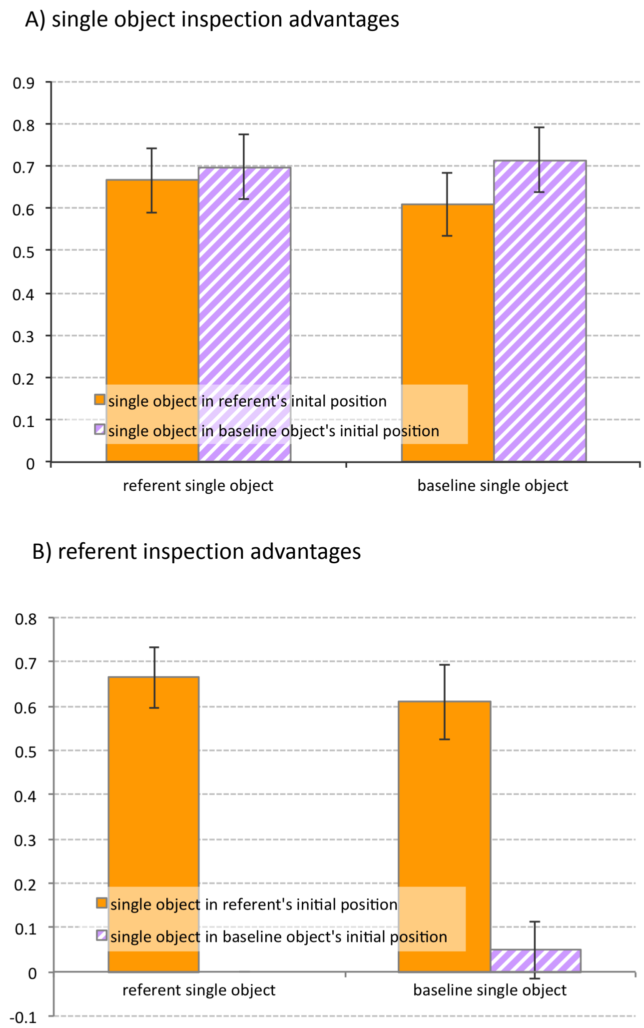

Figure 4.10: Experiment 7: Average single-object (graph A))and referent-position (graph B)) advantages in probabilities of inspections having been started during the shifted word region, one bar per condition. Error bars indicate 95% confidence intervals for both graphs, following Masson and Loftus (2003) for graph A)

strongly affected only by the single object's position ($F(1,23) = 185.6$, $p < .001$; single-object relatedness: $F(1,23) = 1.6$, $p = .23$; interaction: $F(1,23) = 2.0$, $p = .17$), with many more looks to this position when the single object appeared in it than when it was empty. When the referent's initial position was empty from word onset (because the single object was displayed in the baseline object's initial position), it was inspected during the second before the lexical decision response on 49% of trials when the single object was the referent and on 57% of trials when the single object was the baseline object. The difference between these two conditions was not significant (p = .18).

Speed and accuracy of lexical decision

Lexical decision accuracy was not affected by the treatment factors, with an average of 98%, i.e. 5.9 accurate responses out of six trials per participant and condition (single-object relatedness: $F(1,23) = .2$, $p = .66$; single-object position: $F(1,23) = .8$, $p = .38$; interaction: $F(1,23) = .2$, $p = .70$).

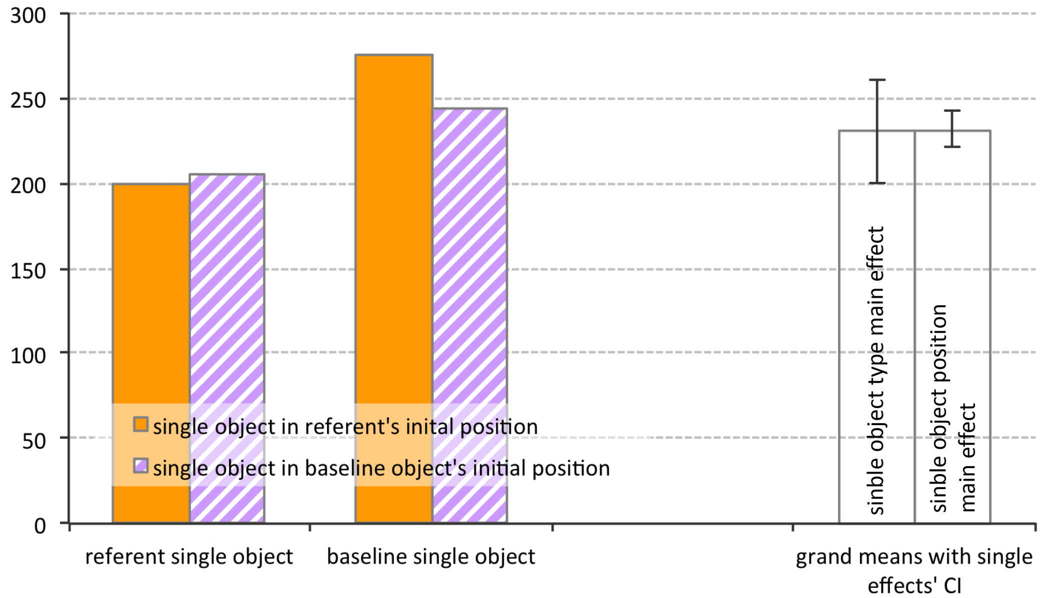

Figure 4.11: Experiment 7: Average lexical decision latencies in msec measured from word end, one bar per condition. Error bars indicating a 99.9% confidence interval for the single-object relatedness main effect and a 95% confidence interval for the single-object position main effect are plotted on separate grand mean bars following Masson and Loftus (2003).

Lexical decision latencies in the entire data set (Figure 4.11 A)) were significantly affected by the single object's relatedness ($F(1,23) = 40.1$, $p < .001$): Lexical decision was faster when the single object was the referent than when it was the baseline object (both $p < .01$). There was also a main effect of the single object's position ($F(1,23) = 5.5$, $p < .05$), carried by lexical decisions being slower when the baseline object was displayed in the referent's initial position than when it was displayed in the baseline object's initial position (p < .05). The interaction of the two factors was not significant ($F(1,23) = 2.7$, $p = .12$)

4.2.3 Discussion

The deployment of visual attention in this experiment followed the visual cue of the display change reliably. Although there were slight imbalances between the, ideally similar, probabilities to have inspected the displayed objects prior to word onset, participants were very likely to shift their gaze to the single object during the spoken word on all conditions. There was also a slight tendency to look at the referent's position after word onset when the single object was the baseline object in its own initial position.

The clear and consistent pattern in lexical decision latencies was that participants were faster to respond to the spoken word when the single object in the display was the word's referent than when it was the baseline object. In addition to this, lexical decision latencies were larger when the baseline object was displayed in the referent's initial position than when it was displayed in its own initial position.

What do these results say about the impact of seeing a referent's visual image and accessing its location on the speed of noun comprehension? Firstly, it is clear that we have found a benefit for word comprehension from seeing the visual image of a referent object while comprehending a spoken noun regardless of its position, i.e. even when it moves to a location previously occupied by another object. Secondly, we need to consider whether we have found evidence that an eye movement to a referent's location provides beneficial information for word comprehension. Lexical decision latencies were not generally faster on those conditions on which participants had to look at the referent's initial position compared to when they had to look at the unrelated baseline object's initial position, so there is no clear evidence for such a benefit. However, we do find evidence for the activation of location-bound information in the difference between the two conditions on which the baseline object was displayed and looked at in its own vs. in the referent's initial position. This is most likely a slow-down for the 'baseline object in the referent's initial position' condition compared to the 'baseline object in its own initial position' condition. It was presumably caused by a conflict between the referent information expected at or associated with that location and the unrelated image actually perceived there. An alternative explanation for the difference in lexical decision latencies between the two conditions with a baseline single object would be that it was not caused by interference when the baseline single object appeared in the referent's initial position, but by a benefit experienced when the baseline single object was seen in its own initial position. Such a benefit could be posited based on the spurious looks to the referent's initial, but empty position on this condition. However, in both Experiment 3 and Experiment 5 of this thesis, looks even to an actual referent object did not produce beneficial effects on word recognition when an unrelated object was highlighted during word comprehension compared to when an unrelated object was highlighted and there never was a referent in the display. For this reason we treat the 'baseline single object in its own initial position' condition of this experiment as a lexical decision baseline, in which neither the referent's image nor access to its location affected word recognition.

4.3 Summary

This chapter presented studies that investigated whether noun-contingent referent looks could enable a faster recognition of the spoken noun via internal information associated to the referent's location in the environment. Experiment 6 found that looks to the *blank location* of a previously-seen referent could not speed up the simultaneous comprehension of a spoken noun. Experiment 7 examined the interplay of the noun-comprehension effects of the *image* provided and the *location* accessed by a noun-contingent referent look. This was accomplished by independently manipulating whether participants looked at a referent or an unrelated object while hearing a spoken noun and whether this object appeared in the same location it had been seen in during the preview period or in the other object's respective position. The identity of the object looked at while hearing a spoken noun had a clear effect on the noun recognition speed: when the referent was seen lexical decision latencies were shorter than when an unrelated object was seen, regardless of whether the referent appeared where participants had seen it before word onset or in a different position. There was also no difference in the noun recognition speed between when the referent was seen in its own vs. the unrelated object's initial position. Interestingly, participants were slower to give a lexical decision response when they looked at an unrelated object in the referent's initial position than when they looked at the unrelated object in the location it was perceived in during the preview period. This can be taken as evidence that the look to the referent's initial position while comprehending the referent's name triggers some associated internal information about the referent, which in turn clashes with the actually encountered unrelated object. Chapter 5 elaborates on conclusions that can be drawn from this result.

5 Conclusions

Noun-contingent eye movements (NCEs) are a well-established behavior in psycholinguistic research: listeners' gaze direction within their visual environment closely mirrors the way in which a spoken utterance refers to entities in that environment. As these listeners have usually had the opportunity to inspect their environment before hearing the utterance and already know the objects they later re-visit with NCEs, it is not immediately evident why they would take the effort to time and direct their eye movement behavior to reflect the referring expressions they comprehend. Visual-world studies utilize people's tendency to make noun-contingent eye movements in order to study the activation of concepts during the processing of spoken language. They have not yet, however, uncovered the motivation behind noun-contingent eye movements. That is why this thesis investigates whether and how noun-contingent eye movements might actually have a positive effect on the comprehension of the nouns that trigger them.

When asking questions about how the processing of one concept affects the processing of another one, the priming and lexical decision paradigms offer an inventory of relevant insights and an adequate methodology to create new ones. For the studies described above, the visual-world, priming and lexical decision methodologies were combined, so that eye movement behavior similar to the one we want to study - noun-contingent eye movements - can be observed at the same time as the speed of comprehension of the spoken noun that triggers them can be assessed. Participants were presented with a display that contained a number of objects (three or four) that they could inspect freely for a certain period of time. After that preview period, they heard a spoken noun, which could be related to one of the displayed objects: it could be the spoken noun's referent, related phonologically or semantically, or it could be entirely unrelated from the spoken noun. Participants' task was to make a lexical decision on the spoken noun (on many filler trials, the spoken stimulus was a non-word). This way the speed of the spoken noun's recognition was observed in relation to the display visual stimuli and together with participants' tendency to make NCEs to the related object.

The first two experiments of this work are pilot studies that use the setup just described, where participants were free to look wherever they liked in response to hearing the spoken noun and one of the displayed objects was either the spoken noun's referent, a phonologically related object (object name and spoken noun shared the same onset), a semantically related object or an unrelated object. The use of these different kinds of objects was motivated by their assumed contrary effects on noun comprehension in a setting like ours: While a referent or a semantically related object could affect noun comprehension beneficially, a phonologically related object could have a detrimental effect on the subsequent processing of a spoken noun. The objective of these two initial experiments was to observe the 'natural' interaction of noun-contingent eye movements and the speed of noun comprehension. While semantically related objects attracted NCEs but did not affect lexical decision latencies, phonologically related objects did slow down lexical decisions when they did not attract looks (in Experiment 1) and had no effect on word comprehension when they were looked at after word onset (Experiment 2). In contrast, the picture observed on those conditions on which the spoken noun's referent was displayed was very clear: The noun's referent consistently attracted NCEs and also made the recognition of the spoken noun faster than when a related or only unrelated objects were displayed.

Experiment 3 has a crucial new methodological feature: at word onset, one object in the display is now highlighted by a visual cue (surrounded by a circle for a very short period of time), so that eye movements during the spoken word are no more directed by word comprehension, but controlled experimentally. This makes it possible to directly observe the *effect* that eye movements during the comprehension of the spoken noun (such as NCEs) have on the speed of comprehension of that noun. The highlighting cue was used to add another factor to the design, in addition to the factor of one object's relatedness to the spoken noun. This new factor determined whether the related object or another, unrelated object was highlighted and thus fixated by participants while they listened to the spoken stimulus. Experiment 3 adds to the evidence found in the pilot studies by proving that looks to a referent object actually *cause* a simultaneously comprehended spoken noun to be recognized faster. This is in comparison to having perceived a referent object during the preview period but not looking at it during the spoken noun's lifetime as well as compared to never having seen the referent. This is an important finding and the central starting point for the further exploration of the mechanisms behind the noun-comprehension benefit from NCEs in Experiments 5 to 7. In Experiment 3, having to look at a phonologically related object during the comprehension of a spoken noun made lexical decisions less accurate, but not significantly slower than having to look at an unrelated object. Taken together with the confirmation of a phonological interference effect with our materials produced in Experiment 4 and with Experiments 1 and 2, we see a rather incoherent picture. This matches the incoherent evidence for phonological interference seen in priming studies (cf. Section 1.3.2). Experiments 5 to 7 therefore focus on noun-contingent *referent* looks and their much more stable benefit for noun recognition. For semantically related objects, the results are more clear: As confirmed by

Experiment 4, an additional priming study with only a single object in the display, our stimuli and setting do not produce semantic facilitation or even interference. So, we cannot conclude that looks towards semantically related objects during word comprehension would not be beneficial. It is conceivable that in a setup where such objects do have an effect on lexical decision in general, looks towards them might also interact with spoken noun comprehension.

Having confirmed in Experiment 3 that referent looks facilitate the simultaneous comprehension of a spoken noun, Experiment 5 tests three hypotheses that detail the timing of this benefit as well as single out the object that causes it. The experiment manipulates whether or not a referent object is shown in the display as well as whether the referent object, empty space or an unrelated object are highlighted at word onset. Triggering looks to empty space at word onset makes it possible to assess, firstly, whether referent looks benefit noun comprehension by preserving a priming effect that originates in the first encounter with the referent object during the preview period (if, when the referent object could be inspected during the preview period, lexical decision is as fast when empty space is fixated during the spoken word as when the referent is fixated from word onset). Secondly, it can be tested whether referent looks benefit noun comprehension simply by avoiding the interference that could arise from looks to unrelated objects (if noun comprehension is slowed down by looking at an unrelated object than by fixating empty space while hearing the spoken noun). These two hypotheses were not confirmed by Experiment 5's results, but instead it could be shown that the benefit from noun-contingent referent looks actually arises only during the comprehension of the spoken noun, and from looking at the referent itself (rather than from not looking at other objects).

The third empirical chapter of this thesis considers an externalist account and examines the word-comprehension benefit of noun-contingent referent looks under the assumption that *spatial pointers* or *indices* are the internal constructs that drive NCEs. On this account, a noun-contingent referent look would cue the access to an internal representation of the referent, which was encoded while the referent was looked at during a trial's preview period. Experiments 6 and 7 search for evidence of such a mechanism. In Experiment 6, participants initially inspect a display containing the referent and two unrelated objects during the preview period. At word onset the screen turns blank except for a circle that highlights either the display location where the referent has been seen before or the location where an unrelated object had been seen. This cue on an otherwise blank screen proved not strong enough to provide access to an internal representation that could affect the word recognition process. Experiment 7 of this work subsequently tests whether the presence of an actual object at the target site of a noun-contingent referent look is necessary in order for a spatially-indexed internal representation to be accessed and to affect noun comprehension. After a preview period identical to Experiment 6's, the display changed at word onset to contain only a single object, which could be the noun's referent or an unrelated object, and it could be located in either the referent's or the unrelated object's preview position. This setting was created to identify the impact of internal information accessed at a noun-contingent eye movement's target location in addition to the impact of the actual visual image provided by such an eye movement. Lexical decisions to the spoken noun were fast whenever the referent object was displayed from word onset on, regardless of its location. Presumably, a benefit from the access to the internal referent representation at the referent's initial location was too weak to speed up lexical decision on top of the already strong benefit from seeing the referent's visual image. Interestingly though, looking at an unrelated object in the referent's initial position caused interference with the lexical decision task compared to when the unrelated object was looked at in its own initial position. This is clearly evidence that an eye movement to the location of a spoken noun's referent activated some kind of an expectation or representation that then created an even bigger interference from the actually encountered unrelated object than that object would have caused in a different location. After some methodological considerations and after relating our findings to the phenomenon of 'real-life' noun-contingent eye movements, a more elaborate account for these results is proposed below.

Our findings about the manner in which eye movements to spoken words' referents can facilitate noun comprehension are based on the observation of two types of eye movement behavior in our studies: In Experiments 1 and 2, participants could freely move their eyes about the display and look at whichever objects they liked during word comprehension. So, whenever they did shift their gaze towards an object related to the spoken word after its onset, they conducted a genuine, voluntary noun-contingent eye movement. In the later experiments of this thesis, in order to be able to examine the causal effect that eye movements have on noun comprehension, participants' visual attention was directed independently of the processes that would normally cause noun-contingent eye movements. This has the side-effect though that there are qualitative differences between the eye movements we induce in our Experiments 3, 5, 6 and 7 and spontaneous noun-contingent referent looks (Experiments 1 and 2).

In natural settings, noun-contingent eye movements are caused by the word being comprehended to a certain degree, by the depicted object being recognized and by a connection between the word and the depicted object being made. Our induced referent looks override such processes by exploiting the attraction of low-level visual (highlighting) cues. We claim, nonetheless, that our results are likely to generalize to real-world NCEs, as we investigated not the processes that trigger NCEs, but the effects of the information provided by them once they have been launched. With regards to timing, our induced eye movements and spontaneous NCEs are rather similar: they start about 200 msec after noun onset and reach their peak from about 500 msec into the word (cf. e.g. Huettig and Altmann (2005) in Section 1.3.1). We manipulate the information which is provided by induced eye movements at the same time during noun comprehension as real-life noun-contingent referent looks would provide similar information. Thus we deem it justified to transfer our findings about the impact of this information on the word comprehension to real-life noun-contingent referent looks and their interaction with word comprehension.

To put our results in perspective, one needs to keep in mind that beyond the beneficial effect of the visual referent information provided by a noun-contingent referent look, those processes that we bypass with our induced referent looks could hypothetically also be beneficial for noun comprehension or even make NCEs superfluous.

Firstly, if one adopts the spatial indexing account for spontaneous NCEs, they are triggered by the activation of the *label* of the referent's spatial index, which is not unlikely to be the referent's name (Zelinsky & Murphy, 2000) and thus directly mapped to the spoken word. One could hypothesize that the activation of this label itself might be sufficient for a benefit on noun comprehension. Under the assumption that the activation of a spatial index through an eye movement to the indexed location also activates that index's label, our Experiment 6 could constitute evidence against this hypothesis: looks to the empty referent location, which should activate the label of the indexed referent, did not speed up lexical decisions to the referent's name. Even if such a benefit from the match of an indexed object's label with the spoken word did exist in the real world though, it could easily co-exist with the referent perception benefit that we observed in our studies.

Secondly, if the comprehension of the spoken noun causes referent looks in the first place, it could be claimed that the triggering of a referent look would take too long for a potential feed-back effect from the NCE back into comprehension, making the NCEs merely an epiphenomenon of the noun comprehension process. This claim seems to hinge critically on the timing of the eye movement in question. The induced eye movements of our experiments did have an effect on the comprehension process, and their timing was similar to the one of real-world NCEs (cf. above): Thus, our referent perception benefit can be taken as evidence that noun-contingent referent looks, even though triggered by the beginnings of word comprehension, can themselves provide information that feeds back into that same comprehension process.

We conducted Experiments 6 and 7 in order to find out whether the eye movement to a referent's position or the fixation on that location would activate internal information that could facilitate the recognition of the referent's name. In Experiment 7, we found interference in lexical decision latencies when an unrelated object was seen in the position that the spoken word's referent had been perceived in earlier, compared to when that same unrelated object was seen in the position that it itself had been perceived in during the preview period. We attribute this interference effect to the fact that an unrelated object is looked at *in the referent's initial position* - on our account, the interference would be due to the activation of the referent's spatial index. There are two ways in which this index's activation could cause interference, which are discussed below and which differ in the way in which participants process the identity of the single object displayed during the word. Both accounts hinge on the idea that a match between the comprehended noun and a spatial index's label (this match does not necessarily need to be on a phonological level, i.e. the label does not need to be the object's name) evokes the implicit hypothesis that the indexed entity could be the name's referent; and that linguistic processing consequently makes itself susceptible to the impact of the visual representation of the indexed entity in order to benefit from the boost that the matching concept of the visual object may give to the processing of the noun.

The first interpretation of Experiment 7's results states that participants, by default, perceived the single object to be the same object as the one displayed in the single object's position before word onset. Inspections towards the unrelated single object in the referent's initial position would be launched with the expectation to encounter the referent, whose actual identity would only become apparent to participants *during* those inspections. Thus, the eye movement towards the single object would activate the referent's spatial pointer, creating an expectation to encounter the referent and making word comprehension susceptible to the influence of the perceived object's influence. When the unrelated object is then seen instead of the referent, it interferes with the recognition of the spoken word. Recall that this interference arises not because the unrelated baseline object is looked at during word comprehension per se, but because the activation of the referent's spatial index through the eye movement to the referent's initial display location makes the comprehension process 'vulnerable' to the impact of that unrelated concept's perception. This would not have happened if the unrelated object's spatial index had been activated together with the visual perception of the unrelated object. This interpretation accounts for the results found for an unrelated single object in Experiment 7, but it is not as straightforward in explaining why there was no difference in lexical decision latencies between the two conditions on which the referent single object was displayed in either its own or the unrelated object's initial position. In principle, one would expect no benefit from seeing a referent when it is located in an unrelated object's location and accessed via that object's spatial index, as the mismatch between that index's label and the spoken word should not permit the referent benefit to affect the spoken word's comprehension. In order to explain our results, one would have to assume that the referent object's impact is so strong that it could affect the recognition of its own name anyways. The next account of our data handles the entirety of Experiment 7's results more gracefully.

The idea that the second interpretation of our results hinges on is that participants perceive the single object to be the same token as one of the previewed objects, which may have moved to a different display location. This is conceivable, as spatial indexing has been shown to be dynamic (Richardson & Kirkham, 2004; Hoover & Richardson, 2008): indexed objects' trajectories around one's visual environments are tracked continuously and the spatial indices' locations are updated accordingly. There would then be two factors that determine whether the activation of the referent's spatial index gets to open up linguistic processing to the influence of visual information: whether or not there is a location in the environment indexed to contain the spoken word's referent, and which kind of relationship there is between the visual information actually found at that location and the spoken word. With regard to these two factors, those two conditions of Experiment 7 where the referent is displayed are virtually similar: in both there is a referent in the display with its spatial pointer pointing to it (although on one condition, when the referent is in the baseline object's initial position, there is also the remaining baseline object's pointer still directed at the referent's location). So, the similar lexical decision benefit on these two conditions can be explained in a simple manner. Similarly, the unrelated single object in the referent's initial position would be the same token as the unrelated object displayed during the preview, but it would now have moved to the referent's initial position. Being forced to look at the unrelated object, no matter where it is located, would activate the *unrelated object's* spatial pointer, there would not be a label-word match and so no comprehension interference would arise. However, if the location previously indexed to contain the *referent's* spatial index is looked at because that is where the unrelated object is currently located, the referent's spatial index is activated as well, and so is the referent's label. The match between this label and the spoken word causes a sensitivity of the linguistic processing for the visual information taken in from the indexed display position. As the information actually perceived is unrelated to the spoken word, however, interference arises (see Figure 5.1 for an illustration of this process). In sum, because it explains our data in a more straightforward way, we adopt this second account for our results on the access of location-bound information through noun-contingent referent looks.

Further experiments will have to be conducted in order to confirm our hypothesis of spatial indices mediating the interaction between a visual object and a spoken noun. Specifically and expanding the experimental setting of Experiment 7, it should be enlightening to independently manipulate the visual image perceived simultaneously with a spoken noun and the presence of spatial indices at that location. The identity of the actual visual image seen during noun comprehension should create a stable effect on noun comprehension times, as has been shown multiple times in this thesis: referents make noun comprehension faster than unrelated objects. Our specific hypothesis would be that this effect arises only when there is a spatial index pointing to

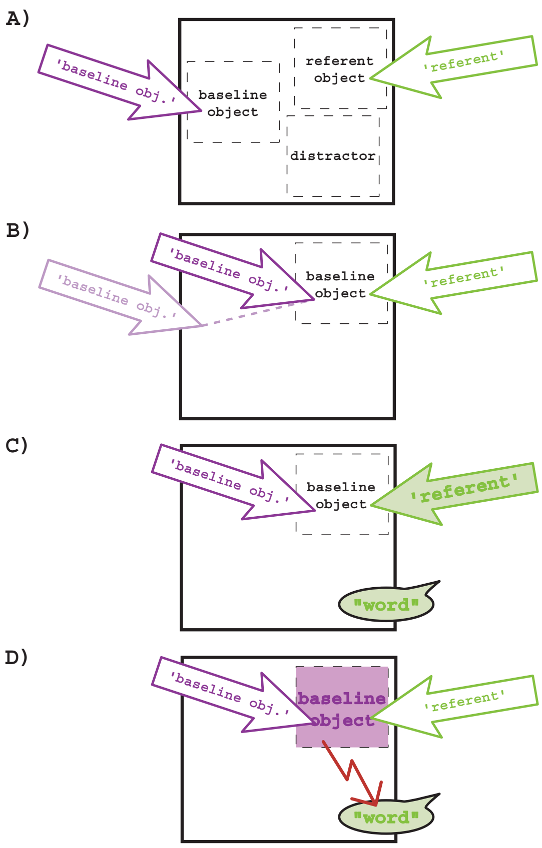

Figure 5.1: An illustration of how we explain the lexical decision interference from the unrelated baseline object appearing in the referent's initial position in Experiment 7: During the preview period, spatial pointers to the unrelated baseline object and the referent are established (A). When only a single object is left on the display, it is perceived to be the same token as one of the objects displayed during the preview; on the condition shown here, the unrelated baseline object is perceived to have moved to the referent's position and spatial pointers are updated accordingly (the referent's index remains unchanged, as the referent has not moved but disappeared) (B). The match between the spoken word and the label of the referent's spatial index is detected (C), and consequently the visual information found at the indexed location is accessed and permitted to interact with the ongoing comprehension of the spoken noun (D). As this visual information does not match the spoken noun in the case depicted here, it interferes with noun comprehension.

the object's location and the match between that pointer's label and the spoken noun enables the interaction of the image with the noun in the first place. This could be confirmed by manipulating the presence of spatial indices associated to the location that the object is perceived in. A relevant experiment would be similar to Experiment 7, with a few changes: Between the preview period and the spoken word's onset, one object in the display (referent or unrelated object) could move to a position not previously inhabited by any other object. At word onset, the single object, which could be the spoken noun's referent or an unrelated object, would appear in this position. In contrast to Experiment 7, it would be clear in this experiment through the depicted motion of objects which spatial index participants access when looking at the single object's position. If our hypotheses hold, the difference in lexical decision latencies between perceiving a referent and an unrelated object should only arise when the object moved to that single position before was the referent object. Through its label's match with the spoken word, only the referent's spatial index would allow the perceived image to affect noun comprehension.

In Experiment 6, on this account, the referent's spatial index would have been active and would have triggered the intake of visual information to feed back into comprehension when the referent's position was highlighted, but the crucial ingredient for an effect on linguistic processing would still have been the actual perception of visual objects, which was not provided in the blank-screen setup of Experiment 6. That is why we found no effect of inspection location on word recognition in that experiment. There is also empirical support for the validity of our findings in Experiment 6: In a similar setting, Hollingworth (2007) finds no location benefit for visual object memory when the surrounding scene disappears between encoding and recall.

To sum up the results of this thesis: Our four initial experiments firmly establish that looking at an object while comprehending its name provides a benefit for that name's comprehension. While our stimuli did not produce any effect of seeing objects semantically related to a spoken noun, phonologically related objects sometimes did and sometimes did not attract visual attention in response to a spoken noun, and did occasionally interfere with the comprehension of that noun. The fifth experiment of this thesis finds that noun-contingent referent looks do not facilitate noun comprehension by avoiding interference that would be caused by seeing unrelated objects or by sustaining a priming effect that stems from seeing the referent before hearing the spoken word, but these looks ease the comprehension of a spoken noun by accessing the referent object simultaneously with the linguistic processing of the noun. The last two experiments explore the role of accessing the referent's location through as part of one's external memory. They find that while merely looking at a referent's empty location during noun comprehension does not affect the speed of comprehension at all, an object in the referent's location does have a strong effect: besides the known and strong benefit from seeing a referent object, an unrelated object in this location can actually have a detrimental effect on noun comprehension (an effect not observed in the previous experiments).

Based on these findings, we propose that spatial indices, specifically the match between an index's label and a comprehended noun, determine the degree of impact that the visual information perceived through a noun-contingent eye movement can have on the comprehension of the spoken noun. A label-word match means that the perceived visual information is free to interact with comprehension, no match means that it cannot. We conclude that spatial pointers do play a crucial role in determining the interaction between linguistic and visual information, by registering whether external visual information could be of interest for the linguistic processor, by mediating the degree of the impact that visual processing has on noun comprehension and by providing access to the visual image of a spoken noun's referent, which can speed up the recognition of that noun's name.

6 Appendix

6.1 Glossary

Critical object: In the experiments described in this thesis, the critical object's processing in relation to the spoken word is of central interest. Its effect on both visual attention and speed of lexical decision is of central interest, and its identity (and thus relation to the spoken word) is manipulated experimentally.

Critical-object-relatedness factor: The factor of the critical object's relatedness to the spoken word experimentally manipulates the identity of the critical object. This object could, for example, be the spoken word's referent, it could be a merely related object or an entirely unrelated object.

Distractor: Distractor objects form a baseline for our visual attention analyses. They are displayed identically on all conditions and can be assumed, due to their unrelatedness to the spoken word and the critical object, not to attract varying amounts of visual attention on the different conditions. Distractors are put to use as visual attention baseline objects in the inspection advantage measures, which are described below.

Highlighted object: The highlighted object is the object or display region that is highlighted via a visual cue at word onset and that should thus be looked at by participants during their perception of the spoken noun. Crucially, the term 'highlighted object' is used orthogonally to other terms like critical object or baseline object. This means that, depending on the condition, the highlighted object can be the critical object, the baseline object or even empty space (in Experiment 5). Consequently, it is a desirable (if paradoxical) result that the likelihood for the highlighted object to be looked at is not affected by the highlighting factor: as the highlighted object is highlighted on every condition, the highlighted object should always be looked at, regardless of any experimental factors.

Highlighting factor: The highlighting factor (in Experiments 3, 5 and 6) determines which object or region in the display is highlighted at word onset (e.g. critical object, baseline object or other).

Inspection advantage: In order to assess whether certain objects in the visual environment receive more visual attention than others, it is necessary to compare their inspection probabilities, which are strongly interdependent measures (as one can only look at one object at a time). Calculating inspection advantages is a way around this problem: The probability to inspect an object that can serve as a visual baseline (here: a distractor) is subtracted from the probability to inspect an object of interest (e.g. a critical object). The resulting inspection-advantage measure is larger than zero if the object of interest received more looks than the visual baseline object, zero if both were looked at equally often and smaller than zero if the visual-attention baseline object received more looks.

Inspection probability: In this thesis, an certain type of inspection probability equals the percentage of trials out of all trials on which that type of inspections took place.

Inspections lasting: This term describes which inspections are counted as having happened within a time region. If an inspection measure is 'Inspections lasting during the time region...' this means that all those inspections that have any temporal overlap with the time region (i.e. that started and/or ended during that time region) are counted for that inspection measure. This procedure is used for inspection measures that describe the likelihood of objects to have been looked at during a certain time region.

Inspections started: This term describes which inspections are counted as having happened within a time region. If an inspection measure is 'Inspections started during the time region...' this means that only those inspections that began during that time region are counted for that inspection measure. This procedure is used for inspection measures that describe shifts of attention.

Shifted time region: Whenever a time region is qualified as *shifted*, this means that 180 msec have been added to its beginning and end. As it takes about 180 msec to plan and execute an eye movement, the shifting of a time region by 180 msec ensures that all those fixations counted within this time region have been launched during the original, non-shifted time region. Time regions are shifted whenever the inspection measures related to this region are analyzed as being affected by experimental stimuli.

Shifted time region before word onset: The shifted time region before word onset begins at trial onset and ends 180 msec after the onset of the spoken word. With the 180-msec saccadic overhead, this region captures the time during which the spoken word could not yet affect inspection probabilities.

Shifted time region of the spoken word: The shifted time region before word onset begins 180 msec after word onset and ends 180 msec after the end of the spoken word. By adding the 180-msec saccadic overhead, this region captures the time during which one can examine the immediate effect of the auditory perception of the spoken word on inspection probabilities. This time region ends at different points in time for every item, as all the spoken words differ in their length.

Simple inspection measures: In contrast to inspection advantage measures (see above), simple inspection measures express the absolute probability that a certain object was looked at, irrespective of the amount of looks received by other objects in the display.

Single object: Term used in Experiment 7, analogous to 'highlighted object' (see above).

Time region of the second before the lexical decision: The time region of the second before the lexical decision begins 1000 msec before that trial's lexical decision was recorded. It captures the time during which participants' viewing behavior could have an effect on their lexical decision. This time region has a constant length, but occurs at a different point in time relative to word onset for every trial.

Unrelated baseline object: In order to asses the effect of looks to a certain object in the display (e.g. of the critical object) on the comprehension of a spoken noun, one needs to be able to compare this effect to a neutral lexical decision baseline. In our experiments, this baseline is provided by looks to the unrelated baseline object, which is unrelated to the spoken noun and remains unchanged across all conditions of an experiment.

6.2 Experimental stimuli

Table 6.1: Experiment 1: Experimental items, part 1. Every experimental item's critical objects.

spoken word / referent critical object	phonologically related critical object	semantically related critical object	unrelated critical object
Auge (*eye*)	Auto (*car*)	Nase (*nose*)	Leiter (*ladder*)
Ballon (*balloon*)	Balkon (*balcony*)	Drachen (*kite*)	Koffer (*suitcase*)
Bart (*beard*)	Bar (*bar*)	Mund (*mouth*)	Tisch (*table*)
Beil (*axe*)	Bein (*leg*)	Dolch (*dagger*)	Schwamm (*sponge*)
Bluse (*blouse*)	Blume (*flower*)	Kleid (*dress*)	Treppe (*staircase*)
Dreirad (*tricycle*)	Dreieck (*triangle*)	Skateboard (*skateboard*)	Steckdose (*outlet*)
Fernglas (*binoculars*)	Fernseher (*television*)	Lupe (*magnifying glass*)	Sichel (*sickle*)
Ferse (*heel*)	Fernrohr (*spy-glass*)	Schulter (*shoulder*)	Trichter (*funnel*)
Flagge (*flag*)	Flasche (*bottle*)	Wappen (*coat of arms*)	Schlüssel (*key*)
Geschirr (*dishes*)	Geschenk (*present*)	Besteck (*cutlery*)	Radio (*radio*)
Gewehr (*gun*)	Gewicht (*balance weight*)	Bogen (*bow*)	Teller (*plate*)
Harke (*rake*)	Harfe (*harp*)	Schaufel (*shovel*)	Puzzle (*puzzle*)
Helm (*helmet*)	Heft (*notebook*)	Schild (*shield*)	Brot (*bread*)
Hose (*trousers*)	Honig (*honey*)	Rock (*skirt*)	Stempel (*postmark*)
Kanone (*cannon*)	Kalender (*calendar*)	Dynamit (*dynamite*)	Briefmarke (*stamp*)
Kastanie (*chestnut*)	Kassette (*cassett*)	Eichel (*acorn*)	Mülltonne (*trash can*)
Kirsche (*cherry*)	Kirche (*church*)	Birne (*pear*)	Sattel (*saddle*)
Münze (*coin*)	Mütze (*cap*)	Geldschein (*banknote*)	Schnabel (*beak*)
Nadel (*needle*)	Nase (*nose*)	Schere (*scissors*)	Lampe (*lamp*)
Nagel (*nail*)	Nase (*nose*)	Hammer (*hammer*)	Teppich (*carpet*)
Palme (*palm*)	Panzer (*tank*)	Kaktus (*cactus*)	Schleife (*bow*)
Schilf (*reed*)	Schild (*shield*)	Hanf (*hemp*)	Hut (*hat*)
Sparschwein (*piggy bank*)	Spargel (*asparagus*)	Geldschein (*banknote*)	Kanu (*canoo*)
Strandkorb (*beach chair*)	Streichholz (*match*)	Sandburg (*sand castle*)	Gabel (*fork*)
Tasse (*cup*)	Tasche (*bag*)	Schale (*bowl*)	Vulkan (*vulcano*)
Trommel (*drum*)	Tropfen (*drop*)	Flöte (*flute*)	Besen (*broom*)
Windel (*diaper*)	Windrad (*wind engine*)	Schnuller (*pacifier*)	Ufo (*UFO*)
Wolle (*wool*)	Wolke (*cloud*)	Faden (*thread*)	Fahrrad (*bike*)
Zirkel (*compass*)	Zirkus (*circus*)	Bleistift (*pencil*)	Krone (*crown*)

Table 6.2: Experiment 1: Experimental items, part 2. Every experimental item's distractors, shown together with their respective referent critical object (which also appears in the first column of the Table 6.1).

spoken word / referent critical object	**distractor 1**	**distractor 2**	**distractor 3**
Auge (*eye*)	Tür (*door*)	Brief (*letter*)	Glas (*glass*)
Ballon (*balloon*)	Turm (*tower*)	Strohhalm (*straw*)	Gehirn (*brain*)
Bart (*beard*)	Nest (*nest*)	Fotoapparat (*camera*)	Spaten (*spade*)
Beil (*axe*)	Laster (*lorry*)	Taschenrechner (*calculator*)	Herd (*stove*)
Bluse (*blouse*)	Pokal (*cup*)	Schreibmaschine (*typewriter*)	Gips (*cast*)
Dreirad (*tricycle*)	Efeu (*ivy*)	Apfel (*apple*)	Pfeife (*pipe*)
Fernglas (*binoculars*)	Schirm (*umbrella*)	Topf (*pot*)	Keule (*club*)
Ferse (*heel*)	Serviette (*napkin*)	Diamant (*diamond*)	Karussell (*caroussel*)
Flagge (*flag*)	Muschel (*shell*)	Glocke (*bell*)	Krawatte (*tie*)
Geschirr (*dishes*)	Pistole (*pistol*)	Schlitten (*sled*)	Blatt (*leaf*)
Gewehr (*gun*)	Mond (*moon*)	Kragen (*collar*)	Sessel (*armchair*)
Harke (*rake*)	Knopf (*button*)	Zeh (*toe*)	Tandem (*tandem*)
Helm (*helmet*)	Maske (*mask*)	Kissen (*pillow*)	Rose (*rose*)
Hose (*trousers*)	Taschenlampe (*torch*)	Bombe (*bomb*)	Kette (*chain*)
Kanone (*cannon*)	Thermometer (*thermometer*)	Erdbeere (*strawberry*)	Spraydose (*spray can*)
Kastanie (*chestnut*)	Gürtel (*belt*)	Löffel (*spoon*)	Portemonnaie (*wallet*)
Kirsche (*cherry*)	Feuerzeug (*lighter*)	Bikini (*bikini*)	Schraube (*screw*)
Münze (*coin*)	Tunnel (*tunnel*)	Klee (*clover*)	Hängematte (*hammock*)
Nadel (*needle*)	Welle (*wave*)	Rennauto (*racing car*)	Zitrone (*lemon*)
Nagel (*nail*)	Kopfhörer (*earphones*)	Fußball (*football*)	Insel (*island*)
Palme (*palm*)	Ei (*egg*)	Tennisschläger (*tennis racket*)	Laterne (*lantern*)
Schilf (*reed*)	Kerze (*candle*)	Teddybär (*teddy bear*)	Büroklammer (*paper clip*)
Sparschwein (*piggy bank*)	Daumen (*thumb*)	Kamm (*comb*)	Ofen (*oven*)
Strandkorb (*beach chair*)	Rasierklinge (*razorblade*)	Locher (*hole punch*)	Nähmaschine (*sewing machine*)
Tasse (*cup*)	Zelt (*tent*)	Seil (*rope*)	Ohr (*ear*)
Trommel (*drum*)	Schaukelstuhl (*rocking chair*)	Grab (*grave*)	Fächer (*fan*)
Windel (*diaper*)	Handschuh (*glove*)	Erdnuß (*peanug*)	Badewanne (*bathtub*)
Wolle (*wool*)	Schlagzeug (*drum set*)	Brunnen (*well*)	Zwiebel (*onion*)
Zirkel (*compass*)	Pilz (*mushroom*)	Aschenbecher (*ashtray*)	Wiege (*cradle*)

Table 6.3: Experiment 2: Experimental items, part 1. Every experimental item's critical objects.

spoken word / referent critical object	**phonologically related critical object**	**semantically related critical object**	**unrelated critical object**
Anorak (*anorak*)	Ananas (*pineapple*)	Overall (*overalls*)	Trompete (*trumpet*)
Ballon (*balloon*)	Balkon (*balcony*)	Drachen (*kite*)	Koffer (*suitcase*)
Bart (*beard*)	Bar (*bar*)	Mund (*mouth*)	Tisch (*table*)
Beil (*axe*)	Bein (*leg*)	Dolch (*dagger*)	Schwamm (*sponge*)
Bluse (*blouse*)	Blume (*flower*)	Kleid (*dress*)	Treppe (*staircase*)
Dreirad (*tricycle*)	Dreieck (*triangle*)	Skateboard (*skateboard*)	Steckdose (*outlet*)
Fernglas (*binoculars*)	Fernseher (*television*)	Lupe (*magnifying glass*)	Sichel (*sickle*)
Ferse (*heel*)	Fernrohr *spy-glass*	Schulter (*shoulder*)	Trichter (*funnel*)
Geschirr (*dishes*)	Geschenk (*present*)	Besteck (*cutlery*)	Radio (*radio*)
Gewehr (*gun*)	Gewicht (*balance weight*)	Bogen (*bow*)	Teller (*plate*)
Harke (*rake*)	Harfe (*harp*)	Schaufel (*shovel*)	Puzzle (*puzzle*)
Helm (*helmet*)	Heft (*notebook*)	Schild (*shield*)	Brot (*bread*)
Kanone (*cannon*)	Kalender (*calendar*)	Dynamit (*dynamite*)	Briefmarke (*stamp*)
Kirsche (*cherry*)	Kirche (*church*)	Birne (*pear*)	Sattel (*saddle*)
Münze (*coin*)	Mütze (*cap*)	Geldschein (*banknote*)	Schnabel (*beak*)
Nadel (*needle*)	Nase (*nose*)	Schere (*scissors*)	Lampe (*lamp*)
Nagel (*nail*)	Nase (*nose*)	Hammer (*hammer*)	Teppich (*carpet*)
Palme (*palm*)	Panzer (*tank*)	Kaktus (*cactus*)	Schleife (*bow*)
Schilf (*reed*)	Schild (*shield*)	Hanf (*hemp*)	Hut (*hat*)
Sparschwein (*piggy bank*)	Spargel (*asparagus*)	Geldschein (*banknote*)	Kanu (*canoo*)
Strandkorb (*beach chair*)	Streichholz (*match*)	Sandburg (*sand castle*)	Gabel (*fork*)
Trommel (*drum*)	Tropfen (*drop*)	Flöte (*flute*)	Besen (*broom*)
Windel (*diaper*)	Windrad (*wind engine*)	Schnuller (*pacifier*)	Ufo (*UFO*)
Wolle (*wool*)	Wolke (*cloud*)	Faden (*thread*)	Fahrrad (*bike*)
Zirkel (*compass*)	Zirkus (*circus*)	Bleistift (*pencil*)	Krone (*crown*)
Auge (*eye*)	Auto (*car*)	Nase (*nose*)	Leiter (*ladder*)
Flagge (*flag*)	Flasche (*bottle*)	Wappen (*coat of arms*)	Schlüssel (*key*)
Kastanie (*chestnut*)	Kassette (*cassett*)	Eichel (*acorn*)	Mülltonne (*trash can*)

Table 6.4: Experiment 2: Experimental items, part 2. Every experimental item's distractors, shown together with their respective referent critical object (which also appears in the first column of the Table 6.3).

spoken word / referent critical object	**distractor 1**	**distractor 2**
Anorak (*anorak*)	Kaffeefilter (*coffee filter*)	Blitz (*lightning*)
Ballon (*balloon*)	Turm (*tower*)	Strohhalm (*straw*)
Bart (*beard*)	Nest (*nest*)	Fotoapparat (*camera*)
Beil (*axe*)	Laster (*lorry*)	Taschenrechner (*calculator*)
Bluse (*blouse*)	Pokal (*cup*)	Schreibmaschine (*typewriter*)
Dreirad (*tricycle*)	Efeu (*ivy*)	Apfel (*apple*)
Fernglas (*binoculars*)	Schirm (*umbrella*)	Topf (*pot*)
Ferse (*heel*)	Serviette (*napkin*)	Diamant (*diamond*)
Geschirr (*dishes*)	Pistole (*pistol*)	Schlitten (*sled*)
Gewehr (*gun*)	Mond (*moon*)	Kragen (*collar*)
Harke (*rake*)	Knopf (*button*)	Zeh (*toe*)
Helm (*helmet*)	Maske (*mask*)	Kissen (*pillow*)
Kanone (*cannon*)	Thermometer (*thermometer*)	Erdbeere (*strawberry*)
Kirsche (*cherry*)	Feuerzeug (*lighter*)	Bikini (*bikini*)
Münze (*coin*)	Tunnel (*tunnel*)	Klee (*clover*)
Nadel (*needle*)	Welle (*wave*)	Rennauto (*racing car*)
Nagel (*nail*)	Kopfhörer (*earphones*)	Fußball (*football*)
Palme (*palm*)	Ei (*egg*)	Tennisschläger (*tennis racket*)
Schilf (*reed*)	Kerze (*candle*)	Teddybär (*teddy bear*)
Sparschwein (*piggy bank*)	Daumen (*thumb*)	Kamm (*comb*)
Strandkorb (*beach chair*)	Rasierklinge (*razorblade*)	Locher (*hole punch*)
Trommel (*drum*)	Schaukelstuhl (*rocking chair*)	Grab (*grave*)
Windel (*diaper*)	Handschuh (*glove*)	Erdnuß (*peanug*)
Wolle (*wool*)	Schlagzeug (*drum set*)	Brunnen (*well*)
Zirkel (*compass*)	Pilz (*mushroom*)	Aschenbecher (*ashtray*)
Auge (*eye*)	Tür (*door*)	Brief (*letter*)
Flagge (*flag*)	Muschel (*shell*)	Glocke (*bell*)
Kastanie (*chestnut*)	Gürtel (*belt*)	Löffel (*spoon*)

Table 6.5: Experiment 3: Experimental items, part 1. Every experimental item's critical objects.

spoken word / referent critical object	**phonologically related critical object**	**semantically related critical object**	**unrelated critical object**
Anorak (*anorak*)	Ananas (*pineapple*)	Overall (*overalls*)	Trompete (*trumpet*)
Ballon (*balloon*)	Balkon (*balcony*)	Drachen (*kite*)	Koffer (*suitcase*)
Beil (*axe*)	Bein (*leg*)	Dolch (*dagger*)	Schwamm (*sponge*)
Bluse (*blouse*)	Blume (*flower*)	Kleid (*dress*)	Treppe (*staircase*)
Dreirad (*tricycle*)	Dreieck (*triangle*)	Skateboard (*skateboard*)	Steckdose (*outlet*)
Fernglas (*binoculars*)	Fernseher (*television*)	Lupe (*magnifying glass*)	Sichel (*sickle*)
Geschirr (*dishes*)	Geschenk (*present*)	Besteck (*cutlery*)	Radio (*radio*)
Gewehr (*gun*)	Gewicht (*balance weight*)	Bogen (*bow*)	Teller (*plate*)
Harke (*rake*)	Harfe (*harp*)	Schaufel (*shovel*)	Puzzle (*puzzle*)
Helm (*helmet*)	Heft (*notebook*)	Schild (*shield*)	Brot (*bread*)
Kanone (*cannon*)	Kalender (*calendar*)	Dynamit (*dynamite*)	Briefmarke (*stamp*)
Kirsche (*cherry*)	Kirche (*church*)	Birne (*pear*)	Sattel (*saddle*)
Nadel (*needle*)	Nase (*nose*)	Schere (*scissors*)	Lampe (*lamp*)
Nagel (*nail*)	Nase (*nose*)	Hammer (*hammer*)	Teppich (*carpet*)
Palme (*palm*)	Panzer (*tank*)	Kaktus (*cactus*)	Schleife (*bow*)
Sparschwein (*piggy bank*)	Spargel (*asparagus*)	Geldschein (*banknote*)	Kanu (*canoo*)
Strandkorb (*beach chair*)	Streichholz (*match*)	Sandburg (*sand castle*)	Gabel (*fork*)
Trommel (*drum*)	Tropfen (*drop*)	Flöte (*flute*)	Besen (*broom*)
Windel (*diaper*)	Windrad (*wind engine*)	Schnuller (*pacifier*)	Ufo (*UFO*)
Wolle (*wool*)	Wolke (*cloud*)	Faden (*thread*)	Fahrrad (*bike*)
Zirkel (*compass*)	Zirkus (*circus*)	Bleistift (*pencil*)	Krone (*crown*)
Auge (*eye*)	Auto (*car*)	Nase (*nose*)	Leiter (*ladder*)
Flagge (*flag*)	Flasche (*bottle*)	Wappen (*coat of arms*)	Schlüssel (*key*)
Kastanie (*chestnut*)	Kassette (*cassett*)	Eichel (*acorn*)	Mülltonne (*trash can*)

Table 6.6: Experiment 3: Experimental items, part 2. Every experimental item's baseline object and distractor, shown together with their respective referent critical object (which also appears in the first column of the Table 6.5).

spoken word / referent critical object	**baseline object**	**distractor**
Anorak (*anorak*)	Kaffeefilter (*coffee filter*)	Blitz (*lightning*)
Ballon (*balloon*)	Turm (*tower*)	Strohhalm (*straw*)
Beil (*axe*)	Laster (*lorry*)	Taschenrechner (*calculator*)
Bluse (*blouse*)	Pokal (*cup*)	Schreibmaschine (*typewriter*)
Dreirad (*tricycle*)	Efeu (*ivy*)	Apfel (*apple*)
Fernglas (*binoculars*)	Schirm (*umbrella*)	Topf (*pot*)
Geschirr (*dishes*)	Pistole (*pistol*)	Schlitten (*sled*)
Gewehr (*gun*)	Mond (*moon*)	Kragen (*collar*)
Harke (*rake*)	Knopf (*button*)	Zeh (*toe*)
Helm (*helmet*)	Maske (*mask*)	Kissen (*pillow*)
Kanone (*cannon*)	Thermometer (*thermometer*)	Erdbeere (*strawberry*)
Kirsche (*cherry*)	Feuerzeug (*lighter*)	Bikini (*bikini*)
Nadel (*needle*)	Welle (*wave*)	Rennauto (*racing car*)
Nagel (*nail*)	Kopfhörer (*earphones*)	Fußball (*football*)
Palme (*palm*)	Ei (*egg*)	Tennisschläger (*tennis racket*)
Sparschwein (*piggy bank*)	Daumen (*thumb*)	Kamm (*comb*)
Strandkorb (*beach chair*)	Rasierklinge (*razorblade*)	Locher (*hole punch*)
Trommel (*drum*)	Schaukelstuhl (*rocking chair*)	Grab (*grave*)
Windel (*diaper*)	Handschuh (*glove*)	Erdnuß (*peanug*)
Wolle (*wool*)	Schlagzeug (*drum set*)	Brunnen (*well*)
Zirkel (*compass*)	Pilz (*mushroom*)	Aschenbecher (*ashtray*)
Auge (*eye*)	Tür (*door*)	Brief (*letter*)
Flagge (*flag*)	Muschel (*shell*)	Glocke (*bell*)
Kastanie (*chestnut*)	Gürtel (*belt*)	Löffel (*spoon*)

Table 6.7: Experiment 4: Experimental items, part 1. Every experimental item's referent, phonologically related and semantically related critical object.

spoken word / referent critical object	**phonologically related critical object**	**semantically related critical object**
Anorak (*anorak*)	Ananas (*pineapple*)	Overall (*overalls*)
Ballon (*balloon*)	Balkon (*balcony*)	Drachen (*kite*)
Bart (*beard*)	Bar (*bar*)	Mund (*mouth*)
Beil (*axe*)	Bein (*leg*)	Dolch (*dagger*)
Bluse (*blouse*)	Blume (*flower*)	Kleid (*dress*)
Dreirad (*tricycle*)	Dreieck (*triangle*)	Skateboard (*skateboard*)
Fernglas (*binoculars*)	Fernseher (*television*)	Lupe (*magnifying glass*)
Ferse (*heel*)	Fernrohr *spy-glass*	Schulter (*shoulder*)
Geschirr (*dishes*)	Geschenk (*present*)	Besteck (*cutlery*)
Gewehr (*gun*)	Gewicht (*balance weight*)	Bogen (*bow*)
Harke (*rake*)	Harfe (*harp*)	Schaufel (*shovel*)
Helm (*helmet*)	Heft (*notebook*)	Schild (*shield*)
Kanone (*cannon*)	Kalender (*calendar*)	Dynamit (*dynamite*)
Kirsche (*cherry*)	Kirche (*church*)	Birne (*pear*)
Münze (*coin*)	Mütze (*hat*)	Geldschein (*banknote*)
Nadel (*needle*)	Nase (*nose*)	Schere (*scissors*)
Nagel (*nail*)	Nase (*nose*)	Hammer (*hammer*)
Palme (*palm*)	Panzer (*tank*)	Kaktus (*cactus*)
Schilf (*reed*)	Schild (*shield*)	Hanf (*hemp*)
Sparschwein (*piggy bank*)	Spargel (*asparagus*)	Geldschein (*banknote*)
Strandkorb (*beach chair*)	Streichholz (*match*)	Sandburg (*sand castle*)
Trommel (*drum*)	Tropfen (*drop*)	Flöte (*flute*)
Windel (*diaper*)	Windrad (*wind engine*)	Schnuller (*pacifier*)
Wolle (*wool*)	Wolke (*cloud*)	Faden (*thread*)
Zirkel (*compass*)	Zirkus (*circus*)	Bleistift (*pencil*)
Auge (*eye*)	Auto (*car*)	Nase (*nose*)
Flagge (*flag*)	Flasche (*bottle*)	Wappen (*coat of arms*)
Kastanie (*chestnut*)	Kassette (*cassett*)	Eichel (*acorn*)

Table 6.8: Experiment 4: Experimental items, part 2. Every experimental item's unrelated critical object and baseline object, shown together with their respective referent critical object (which also appears in the first column of the Table 6.7).

spoken word / referent critical object	**unrelated critical object**	**baseline object**
Anorak (*anorak*)	Trompete (*trumpet*)	Kaffeefilter (*coffee filter*)
Ballon (*balloon*)	Koffer (*suitcase*)	Strohhalm (*straw*)
Bart (*beard*)	Tisch (*table*)	Nest (*nest*)
Beil (*axe*)	Schwamm (*sponge*)	Laster (*lorry*)
Bluse (*blouse*)	Treppe (*staircase*)	Schreibmaschine (*typewriter*)
Dreirad (*tricycle*)	Steckdose (*outlet*)	Efeu (*ivy*)
Fernglas (*binoculars*)	Sichel (*sickle*)	Topf (*pot*)
Ferse (*heel*)	Trichter (*funnel*)	Serviette (*napkin*)
Geschirr (*dishes*)	Radio (*radio*)	Pistole (*pistol*)
Gewehr (*gun*)	Teller (*plate*)	Kragen (*collar*)
Harke (*rake*)	Puzzle (*puzzle*)	Knopf (*button*)
Helm (*helmet*)	Brot (*bread*)	Kissen (*pillow*)
Kanone (*cannon*)	Briefmarke (*stamp*)	Thermometer (*thermometer*)
Kirsche (*cherry*)	Sattel (*saddle*)	Bikini (*bikini*)
Münze (*coin*)	Schnabel (*beak*)	Tunnel (*tunnel*)
Nadel (*needle*)	Lampe (*lamp*)	Welle (*wave*)
Nagel (*nail*)	Teppich (*carpet*)	Fussball (*football*)
Palme (*palm*)	Schleife (*bow*)	Ei (*egg*)
Schilf (*reed*)	Hut (*hat*)	Kerze (*candle*)
Sparschwein (*piggy bank*)	Kanu (*canoo*)	Kamm (*comb*)
Strandkorb (*beach chair*)	Gabel (*fork*)	Rasierklinge (*razorblade*)
Trommel (*drum*)	Besen (*broom*)	Grab (*grave*)
Windel (*diaper*)	Ufo (*UFO*)	Handschuh (*glove*)
Wolle (*wool*)	Fahrrad (*bike*)	Brunnen (*well*)
Zirkel (*compass*)	Krone (*crown*)	Pilz (*mushroom*)
Auge (*eye*)	Leiter (*ladder*)	Brief (*letter*)
Flagge (*flag*)	Schlüssel (*key*)	Muschel (*shell*)
Kastanie (*chestnut*)	Mülltonne (*trash can*)	Löffel (*spoon*)

Table 6.9: Experiment 5: All the experimental items of Experiment 5.

spoken word / referent critical object	**unrelated critical object**	**baseline object**	**distractor**
Anorak (*anorak*)	Trompete (*trumpet*)	Wiege (*cradle*)	Gießkanne (*watering can*)
Gebiss (*set of teeth*)	Beil (*axe*)	Schneeflocke (*snowflake*)	Hängematte (*hammock*)
Bluse (*blouse*)	Treppe (*staircase*)	Fernrohr *spy-glass*	Pokal (*cup*)
Dreirad (*tricycle*)	Steckdose (*outlet*)	Krawatte (*tie*)	Walnuß (*walnut*)
Fernglas (*binoculars*)	Klee (*clover*)	Ofen (*oven*)	Säule (*pillar*)
Geschirr (*dishes*)	Lupe (*magnifying glass*)	Nest (*nest*)	Schuh (*shoe*)
Gewehr (*gun*)	Würfel (*dice*)	Teller (*plate*)	Schere (*scissors*)
Wolke (*cloud*)	Puzzle (*puzzle*)	Zwiebel (*onion*)	Knopf (*button*)
Helm (*helmet*)	Brot (*bread*)	Auto (*car*)	Geige (*violin*)
Kanone (*cannon*)	Briefmarke (*stamp*)	Erdbeere (*strawberry*)	Iglu (*igloo*)
Kirsche (*cherry*)	Sattel (*saddle*)	Gehirn (*brain*)	Schwamm (*sponge*)
Lampe (*lamp*)	Nadel (*needle*)	Butter (*butter*)	Welle (*wave*)
Nagel (*nail*)	Stempel (*postmark*)	Fächer (*fan*)	Kopfhörer (*earphones*)
Palme (*palm*)	Schleife (*bow*)	Laptop (*laptop*)	Zirkus (*circus*)
Strandkorb (*beach chair*)	Overall (*overalls*)	Löwenzahn (*dandelion*)	Fahrrad (*bike*)
Trommel (*drum*)	Besen (*broom*)	Kissen (*pillow*)	Flasche (*bottle*)
Windel (*diaper*)	Ufo (*UFO*)	Rollo (*blind*)	Kerze (*candle*)
Wolle (*wool*)	Tandem (*tandem*)	Birne (*pear*)	Dynamit (*dynamite*)
Zirkel (*compass*)	Krone (*crown*)	Seil (*rope*)	Pilz (*mushroom*)
Blatt (*leaf*)	Wurst (*sausage*)	Ohr (*ear*)	Leiter (*ladder*)
Flagge (*flag*)	Harfe (*harp*)	Armbanduhr (*watch*)	Banane (*banana*)
Kastanie (*chestnut*)	Mülltonne (*trash can*)	Sanduhr (*hourglass*)	Gürtel (*belt*)
Tasse (*cup*)	Pflaster *band-aid*	Vulkan (*vulcano*)	Schraube (*screw*)
Tisch (*table*)	Bart (*beard*)	Zelt (*tent*)	Panzer (*tank*)
Lenkrad (*steering wheel*)	Socke (*sock*)	Flöte (*flute*)	Schaufel (*shovel*)
Kanne (*pitcher*)	Brille (*glasses*)	Anker (*anchor*)	Garage (*garage*)
Zahn (*tooth*)	Blume (*flower*)	Schrank (*closet*)	Magnet (*magnet*)
Sessel (*armchair*)	Weste (*vest*)	Korb (*basket*)	Tablette (*tablet*)
Geschenk (*present*)	Schubkarre (*wheelbarrow*)	Mauer (*wall*)	Zigarre (*cigar*)
Golfschläger (*golf club*)	Lippenstift (*lipstick*)	Kompaß (*compass*)	Teddybär (*teddy bear*)
Tabelle (*table*)	Kaktus (*cactus*)	Feuerzeug (*lighter*)	Hamburger (*hamburger*)
Saxophon (*saxophone*)	Flugzeug (*plane*)	Kamm (*comb*)	Wappen (*coat of arms*)
Grab (*grave*)	Ei (*egg*)	Bus (*bus*)	Kassette (*cassett*)
Federball (*shuttlecock*)	Lutscher (*lollipop*)	Hufeisen (*horseshoe*)	Mikrophon (*microphone*)
Streichholz (*match*)	Zitrone (*lemon*)	Bettwäsche *bed-linen*	Laster (*lorry*)
Buch (*book*)	Schild (*shield*)	Tennisball (*tennis ball*)	Insel (*island*)
Pinzette (*tweezers*)	Sonnenblume (*sunflower*)	Bar (*bar*)	Hantel (*barbell*)
Badewanne (*bathtub*)	Maske (*mask*)	Leuchtturm (*lightouse*)	Rose (*rose*)
Heft (*notebook*)	Bank (*bank*)	Klavier (*piano*)	Eichel (*acorn*)
Kiwi (*kiwi*)	Jackett (*jacket*)	Balkon (*balcony*)	Traktor (*tractor*)
Sofa (*sofa*)	Glocke (*bell*)	Apfel (*apple*)	Schlüssel (*key*)
Brief (*letter*)	Eis (*ice cream*)	Gips (*cast*)	Turm (*tower*)
Gewicht (*balance weight*)	Becher (*cup*)	Rakete (*rocket*)	Tür (*door*)
Drachen (*kite*)	Spargel (*asparagus*)	Taschenrechner (*calculator*)	Bikini (*bikini*)
Skateboard (*skateboard*)	Goldbarren (*gold bar*)	Tornado (*tornado*)	Vase (*vase*)
Sparschwein (*piggy bank*)	Locher (*hole punch*)	Wecker (*alarm clock*)	Kette (*chain*)
Schnabel (*beak*)	Windrad (*wind engine*)	Handtuch (*towel*)	Fernseher (*television*)
Ferse (*heel*)	Trichter (*funnel*)	Schaukel (*swing*)	Maßband (*measuring tape*)

Table 6.10: Experiments 6 and 7: All the experimental items of Experiments 6 and 7.

spoken word / referent critical object	**baseline object**	**distractor**
Anorak (*anorak*)	Trompete (*trumpet*)	Kaffeefilter (*coffee filter*)
Ballon (*balloon*)	Koffer (*suitcase*)	Turm (*tower*)
Beil (*axe*)	Schwamm (*sponge*)	Laster (*lorry*)
Bluse (*blouse*)	Treppe (*staircase*)	Pokal (*cup*)
Dreirad (*tricycle*)	Steckdose (*outlet*)	Efeu (*ivy*)
Fernglas (*binoculars*)	Sichel (*sickle*)	Schirm (*umbrella*)
Geschirr (*dishes*)	Radio (*radio*)	Pistole (*pistol*)
Gewehr (*gun*)	Teller (*plate*)	Mond (*moon*)
Harke (*rake*)	Puzzle (*puzzle*)	Knopf (*button*)
Helm (*helmet*)	Brot (*bread*)	Maske (*mask*)
Kanone (*cannon*)	Briefmarke (*stamp*)	Thermometer (*thermometer*)
Kirsche (*cherry*)	Sattel (*saddle*)	Feuerzeug (*lighter*)
Nadel (*needle*)	Lampe (*lamp*)	Welle (*wave*)
Nagel (*nail*)	Teppich (*carpet*)	Kopfhörer (*earphones*)
Palme (*palm*)	Schleife (*bow*)	Ei (*egg*)
Sparschwein (*piggy bank*)	Kanu (*canoo*)	Daumen (*thumb*)
Strandkorb (*beach chair*)	Gabel (*fork*)	Rasierklinge (*razorblade*)
Trommel (*drum*)	Besen (*broom*)	Schaukelstuhl (*rocking chair*)
Windel (*diaper*)	Ufo (*UFO*)	Handschuh (*glove*)
Wolle (*wool*)	Fahrrad (*bike*)	Schlagzeug (*drum set*)
Zirkel (*compass*)	Krone (*crown*)	Pilz (*mushroom*)
Auge (*eye*)	Leiter (*ladder*)	Tür (*door*)
Flagge (*flag*)	Schlüssel (*key*)	Muschel (*shell*)
Kastanie (*chestnut*)	Mülltonne (*trash can*)	Gürtel (*belt*)

References

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

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

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

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

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

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*, 55–71.

Anderson, J. E., & Holcomb, P. J. (1995). Auditory and visual semantic priming using different stimulus onset asynchronies: An event-related brain potential study. *Psychophysiology*, *32*, 177–190.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*, B13–B26.

Arthur Glenberg, J. S., & Robertson, D. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition*, *26*(4), 651–658.

Baayen, H. R. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers*, *1*, 1–45.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic code for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 723–767.

Balota, D. A., & Robert F. Lorch, J. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *12*(3), 336–345.

Boland, J. E. (2005). Visual arguments. *Cognition*, *95*, 237–274.

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.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*, 633–664.

Dahan, D., & Gaskell, G. M. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language*, *57*, 483–501.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5/6), 507–534.

Dahan, D., Tanenhaus, M., & Salverda, A.-P. (2007). The influence of visual processing on phonetically driven saccades in the "visual world" paradigm. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movement research: A window on mind and brain* (pp. 471–486). Oxford: Elsevier.

Dennis Norris, J. M. M., Anne Cutler, & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology*, *53*, 146–193.

Deutscher Wortschatz. (2007). Available from `http://wortschatz.uni-leipzig.de/`

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *TRENDS in Cognitive Sciences*, *12*(11), 405–410.

Forster, K. I., & Masson, M. E. J. (Eds.). (2008). Emerging data analysis [special issue]. *Journal of Memory and Language*, *59*(4).

Goldinger, S. D. (1996). Auditory lexical decision. *Language and Cognitive Processes*, *11*(6), 559–567.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The role of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1211–1238.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, *14*(5), 462–466.

Griffin, Z. M. (2004). Why look? Reasons for eye movements related to language production. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world.* New York: Psychology Press.

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

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes*, *8*(4), 379–411.

Hollingworth, A. (2006). Scene and specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*(1), 58–69.

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 31–47.

Hoover, M. A., & Richardson, D. C. (2008). When facts go down the rabbit hole: Contrasting features and objecthood as indexes to memory. *Cognition*, *108*, 533–542.

Huettig, F., & Altmann, G. T. M. (2004). The online processing of ambiguous and unambiguous words in context: Evidence from head-mounted eye-tracking. In M. Carreiras & C. E. Clifton (Eds.), *The on-line study of sentence comprehension: eyetracking, ERP and beyond.* New York: Psychology Press.

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

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, *15*(8), 985–1018.

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*, 460–482.

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin*, *96*(1), 29–44.

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

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from german and english. *Journal of Psycholinguistic Research*, *32*(1), 37–55.

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

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

Koivisto, M., & Revonsuo, A. (2000). Semantic priming by pictures and words in the cerebral hemispheres. *Cognitive Brain Research*, *10*, 91–98.

Levelt, W. J. M., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*(1), 122–142.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, *28*, 103–174.

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

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, *57*(3), 203–220.

Matsukawa, J., Snodgrass, J. G., & Doniger, G. M. (2005). Conceptual versus perceptual priming in incomplete picture identification. *Journal of Psycholinguistic Research*, *34*(6), 515–540.

McMurray, B., Tanenhaus, M. K., Aslin, R. N., & Spivey, M. J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research*, *32*(1), 77–97.

McQueen, J. M., Dahan, D., & Cutler, A. (2004). Continuity and gradedness in speech processing. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production. differences and similarities.* Berlin, New York: Mouton de Gruyter.

McQueen, J. M., & Huettig, F. (2005). Semantic and phonological priming of auditory lexical decision by pictures and printed words. In *Proceedings of the 11th conference on architectures and mechanisms of language processing.* Ghent, Belgium.

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General*, *132*(2), 228–252.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, *6*(2), 182–189.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 863–883.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, *4*(5), 648–654.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226–254.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*(3), 461–488.

Raaijmakers, J. G. W. (2003). A further look at the "language-as-fixed-effect fallacy". *Canadian Journal of Experimental Psychology*, *57*(3), 141–151.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language*, *41*, 416–426.

Radeau, M., Morais, J., & Segui, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1297–1311.

Richardson, D. C., Altmann, G. T. M., Spivey, M. J., & Hoover, M. A. (2009). Much ado about eye movements to nothing: A response to Ferreira et al.: Taking a new look at looking at nothing. *TRENDS in Cognitive Sciences*, *13*(6), 235–236.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse compehension. *Cognitive Science*, *29*, 1045–1060.

Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General*, *133*(1), 46–62.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*, 269–295.

Salverda, A. P., & Altmann, G. T. M. (2005). Interference of visually-cued eye movements by spoken language. In *Proceedings of the 11th conference on architectures and mechanisms of language processing.* Ghent, Belgium.

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

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition*, *12*(4), 315–328.

Spivey, M. J., Richardson, D. C., & Fitneva, S. (2004). Thinking outside the brain: Spatial indices to visual and linguistic information. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision and action: Eye movements and the visual world.* New York: Psychology Press.

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

Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, *120*, 268–291.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. G. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557–580.

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

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*(5), 379–385.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*(3), 367–392.

Yee, E., & Sedivy, J. (2001). Using eye movements to track the spread of semantic activation during spoken word recognition. In *Proceedings of the 13th annual cuny sentence processing conference.* Philadelphia, USA.

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

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science*, *11*(2), 125–131.