Recovery from misinterpretations during online sentence processing

Lena M Blott[1a], Jennifer M Rodd[1], Fernanda Ferreira[2], and Jane E Warren[1]

[1] Division of Psychology and Language Sciences, University College London, UK

[2] Department of Psychology, University of California, Davis, CA, USA

[a] Corresponding author:

Lena M Blott

Chandler House

2 Wakefield Street

London WC1N 1PF

United Kingdom

Email: lena.blott.12@ucl.ac.uk

Abstract

Misinterpretations during language comprehension are common. The ability to recover from such processing difficulties is therefore crucial for successful day-to-day communication. The present study investigated the outcome of comprehension processes and on-line reading behaviour when misinterpretations occurred. Although group-level effects of reinterpretation on sentence comprehension and on-line processing are of great theoretical interest, individual differences in the recovery from processing difficulty are of particular practical relevance. Even adult readers vary considerably in their “lexical expertise”, their knowledge of word forms and meanings and their experience with written material. We therefore also investigated the effect of individual differences in lexical expertise on processes related to the recovery from misinterpretations. Ninety-six adult participants read “garden-path” sentences in which an ambiguous word was disambiguated towards an unexpected meaning (e.g. *The ball was crowded*), while their eye movements were monitored. A Meaning Coherence Judgement task additionally required them to decide whether or not each sentence made sense. Results suggested that readers did not always engage in reinterpretation processes but instead followed a “good enough” processing strategy. Successful detection of a violation to sentence coherence and associated reinterpretation processes also required additional processing time compared to sentences that did not induce a misinterpretation. Although these reinterpretation-related processing costs were relatively stable across individuals, there was some evidence to suggest that readers with greater lexical expertise benefited from greater sensitivity to the disambiguating information, and were able to flexibly adapt their on-line reading behaviour to recover from misinterpretations more efficiently.

Keywords: misinterpretations, sentence comprehension, vocabulary, print exposure, eye-tracking

Language comprehension is a highly complex task: to understand a sentence, we need to map word forms to meanings, and rapidly access and integrate appropriate lexical representations into a coherent representation of sentence meaning. As we process sentences, we do not wait until the end of a clause or a sentence to start accessing word meanings and build up sentence structure; instead, we use the information we already have to process sentences as we go (e.g. Just & Carpenter, 1980; Marslen-Wilson, 1973, 1975; Traxler & Pickering, 1996; Tyler & Marslen-Wilson, 1977). This incrementality is advantageous for rapid comprehension but comes at a cost – for example, you may have likely initially misinterpreted the sentence in *Figure 1*. To successfully understand sentences, we also need to be able to detect errors in meaning coherence, and trigger appropriate recovery procedures to restore this coherence.

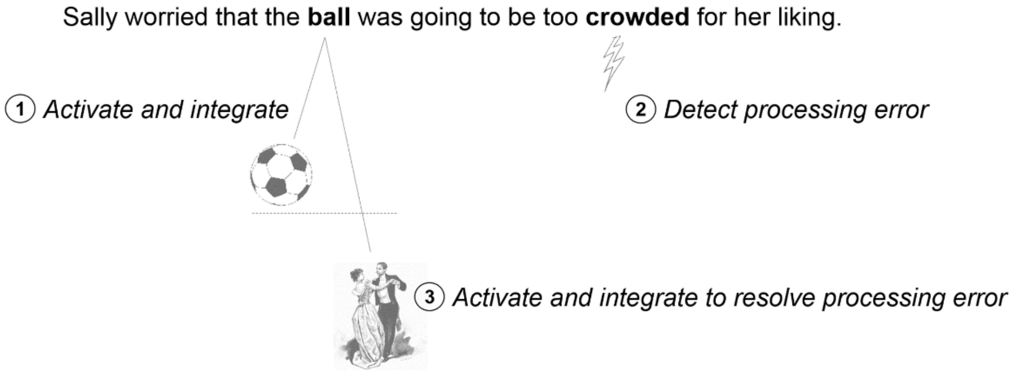

Figure 1. Illustration of reinterpretation processes in a semantic garden-path sentence. A reader will usually initially activate and integrate the dominant (most frequent) meaning of the ambiguous word (1). However, to understand this sentence successfully, she will need to detect a violation to sentence coherence at the disambiguating information (2), and trigger appropriate reinterpretation processes to activate and integrate the intended subordinate meaning of the ambiguous word into the sentence context (3).

Figure 1 provides a prototypical example of a sentence that requires reinterpretation. Although this example has been specifically created to lead you down a metaphorical “garden path” during your initial interpretation, the phenomenon illustrated here is not uncommon in the language we hear and read daily. The “garden path” in this sentence relies on the fact that the word form *ball* maps onto multiple semantic representations: *ball* can refer to a spherical toy or to a formal dancing event. Lexical-semantic ambiguity, or the one-to-many mapping between orthography/phonology and semantics is very common; some estimates put the proportion of ambiguous words among frequently used English words as high as 80% (Rodd et al., 2002, p. 250). Sentences that contain an ambiguous word can therefore be used to investigate processes related to the selection of appropriate word meanings, the detection of coherence violations, and reinterpretation.

In the example in *Figure 1*, the ambiguous word *ball* is disambiguated towards its less expected, subordinate meaning ("formal dancing event") by the word *crowded*. However, the reader is not given any information about the meaning of the ambiguous word before its encounter. The question of whether all potential meanings of an ambiguous word are always activated in parallel or whether only a single meaning receives activation has historically received much attention (e.g. Foss & Jenkins, 1973; Hogaboam & Perfetti, 1975; Holmes, 1979; Onifer & Swinney, 1981; Schvaneveldt, Meyer, & Becker, 1976; Simpson, 1981; Simpson & Krueger, 1991; Swinney, 1979; Tabossi, 1988). Models of lexical ambiguity processing have converged on the view that although all meanings are briefly activated, the processing system rapidly settles on a single meaning (Cairns & Kamerman, 1975; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Twilley & Dixon, 2000). When previous context is uninformative (as is the case in *Figure 1*), the processing system makes that meaning selection using a best-guess default heuristic based on the relative occurrence frequencies of the respective meanings of the ambiguous word (Duffy, Morris, & Rayner, 1988; Gadsby, Arnott, & Copland, 2008; Simpson, 1981; Simpson & Burgess, 1985; Twilley & Dixon, 2000), or on the basis of its most recent encounter with the word form (Rodd et al., 2016; Rodd, Lopez Cutrin, Kirsch, Millar, & Davis, 2013). As the most frequent, or "dominant", meaning of *ball* is the "spherical toy" meaning, and there is no prior disambiguating context available in the example sentence, the processing system will initially select this meaning, and therefore be led down the garden path. This garden-path effect persists until the disambiguating information (*crowded*) is encountered. At this point in the sentence, the system needs to detect a violation to sentence-meaning coherence as the selected meaning of *ball* is incompatible with the context provided by *crowded*. The detection of the coherence violation needs to be followed by reinterpretation procedures that identify the ambiguity in *ball* as the culprit for the comprehension difficulty, access its previously-discarded subordinate meaning (“dancing event”), and integrate it into the sentence context.

The recovery from so-called “garden-path” sentences has previously been associated with processing costs. Readers and listeners generally take longer to process sentences that require reinterpretation compared to sentences that do not contain coherence violations (e.g. Dopkins, 1992; Duffy, Morris, & Rayner, 1988; Foss, Bever, & Silver, 1968; Holmes, Arwas, & Garrett, 1977; Rayner & Duffy, 1986; Rayner & Frazier, 1989; Rayner, Pacht, & Duffy, 1994; Sereno, Pacht, & Rayner, 1992). More specifically, eye-tracking research has found that readers spend longer on sentence regions containing disambiguating information (“crowded” in *Figure 1*) compared to equivalent phrases in unambiguous sentences, indicative of the detection of a coherence violation (Ferreira & Clifton, 1986; Ferreira & Henderson, 1991). Similarly, leftward regressive eye movements out of such disambiguating regions back to previous sentence regions are thought to indicate difficulty in integrating the disambiguating region into prior context. Such physical re-reading is therefore assumed to reflect attempts at recovery from a processing difficulty (Frazier & Rayner, 1987; Pickering & Traxler, 1998). It has often been implicitly assumed that such ambiguity-related processing difficulty is temporary, and that the outcome of language processing operations is a veridical representation of sentence structure and meaning as presented. However, there is compelling evidence to suggest that comprehenders do not always build detailed sentence representations. For example, when reading syntactically ambiguous sentences such as “While Anna dressed the baby played in the crib”, comprehenders often incorrectly assign “the baby” the patient role for the verb “dressed” (Christianson, Hollingworth, Halliwell, & Ferreira, 2001). This and similar findings have led to the development of the theory of “good enough processing”, which assumes that the processing system is often content constructing linguistic representations that are superficial but “good enough” for the job (Ferreira, 2003; Ferreira &

Patson, 2007; Karimi & Ferreira, 2016). A “good enough” processor may therefore have missed the coherence violation in the example sentence, and not engaged in reinterpretation processes.

Much of the evidence for reading behaviour in “garden path” sentences is based on syntactic garden-paths (e.g. “The horse raced past the barn fell”), while lexical-semantic ambiguity has featured more prominently in research on meaning access and selection processes during single-word reading. Although this separation has created the impression that these are two distinct types of ambiguity, it has been argued that syntactic ambiguity – like lexical-semantic ambiguity – ultimately relies on the ambiguity of lexical representations (MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994). Since the resolution of ambiguity, and therefore the recovery from ambiguity-related misinterpretations, relies on access to lexical information, *how* this information is represented may be critical for the success of the required reinterpretation processes. Individual differences in the ability to access rich lexical representations, and to build and maintain elaborated representations of context may affect both the efficiency and the effectiveness with which a comprehender can recover from a misinterpretation (see also Daneman & Carpenter, 1983; Gernsbacher & Faust, 1991; Twilley & Dixon, 2000). Language comprehension during reading is associated with large inter-individual variability, even among adult readers. Although group-level effects of reinterpretation demands on processing are of theoretical interest, the implications of individual differences in reading behaviour and comprehension are arguably of greater practical relevance. For example, the ability to effectively and efficiently recover from misinterpretations may put limits to an individual’s educational attainment, and make day-to-day functioning in a text-reliant society, where the written word has overtaken speech as the primary means of communication, unnecessarily effortful.

In their Reading Systems Framework, Perfetti and Stafura (2014) imply that readers who need to expend excessive resources on lexical access may only have limited resources available for repair processes. According to the Lexical Quality Hypothesis proposed by Perfetti and colleagues (Perfetti, 2007; Perfetti & Hart, 2002), high-quality lexical representations are characterised by well-specified and well-connected information relating to orthographic, phonological, and semantic aspects of a word. The quality of lexical representations is thought to affect integration of words within sentence contexts, such that weak lexical representations may cause disruption to downstream comprehension processes (Perfetti & Stafura, 2014). In line with this idea, successful text comprehension has been strongly linked to lexical knowledge, both in beginning and in more experienced readers (Adlof, Catts, & Little, 2006; Braze, Tabor, Shankweiler, & Mencl, 2007; Cromley & Azevedo, 2007; Guo, Roehrig, & Williams, 2011; Malatesha Joshi, 2005; Perfetti & Hart, 2002; Prat & Just, 2011; Traxler & Tooley, 2007; Tunmer & Chapman, 2012). Adults readers vary considerably in their knowledge of word forms and meanings (e.g. Mainz, Shao, Brysbaert, & Meyer, 2017), and in their experience with written material (Stanovich & West, 1989; Guthrie & Seifert, 1983). Extensive experience processing written texts likely improves the efficiency of word-to-sentence integration processes, and may enhance the knowledge of suitable strategies to resolve processing difficulties (see also arguments about the role of experience made in Farmer et al., 2012; MacDonald & Christiansen, 2002; MacDonald et al., 1994). Readers' experience with words in different contexts also supports the development of high-quality lexical representations (Yee & Thompson-Schill, 2016). This "lexical expertise", readers' accumulated lexical knowledge and experience, may thus influence the ability to recover from the kind of misinterpretation illustrated in *Figure 1*.

While there exists a vast literature on variation in lexical knowledge and its influence on reading processes among beginning readers (and rightfully so), there is a surprising paucity in research on individual differences in lexical expertise and reading processes in adults. Although lexical access and word-to-sentence integration processes are vital to successful reading comprehension and subject to individual differences even among literate adults, very few studies have investigated the role of lexical knowledge during on-line sentence processing in adult readers (but see e.g. Federmeier, McLennan, De Ochoa, & Kutas, 2002). Previous research investigating eye movement behaviour during reading suggests that greater lexical expertise is associated with facilitated word recognition and lexical access (Ashby, Rayner, &

Clifton, 2005), and more efficient incremental sentence-level processing (Luke, Henderson, & Ferreira, 2015; Payne, Gao, Noh, Anderson, & Stine-Morrow, 2012; Taylor & Perfetti, 2016; Wells, Christiansen, Race, Acheson, & Macdonald, 2009). However, there is currently a gap in the literature concerning the role of lexical expertise in comprehension and on-line reading behaviour. An important question is whether lexical knowledge and reading experience affect reinterpretation processes in experienced readers. The aims of the present study were therefore to investigate both the outcome of comprehension processes and on-line reading behaviour in sentences that require reinterpretation, and to explore whether these processes are affected by readers' lexical expertise.

In the present study, we used semantic garden-path sentences like the example in *Figure 1* to investigate reinterpretation processes. These sentences have been shown to reliably introduce a misinterpretation that needs to be resolved for successful comprehension. We used eye-tracking during reading in conjunction with a behavioural task to be able to measure both on-line processing and off-line comprehension of these sentences. Sentence comprehension was measured with a Meaning Coherence Judgement task in which readers decided whether or not a sentence made sense. In addition to semantic garden-paths that contained an ambiguous noun that was disambiguated towards its unexpected meaning later in the sentence, matched unambiguous sentences were created. The Meaning Coherence Judgement task also required the creation of sentences that did not make sense. We therefore also included semantically anomalous sentences that either contained an ambiguous or an unambiguous noun, whose meaning(s) were incompatible with the remainder of the sentence content. During sentence reading, we measured participants' eye movements, and collected data from three main regions-of-interest: the Main Noun region (corresponding to point (1) in *Figure 1*), the Coherence cue region where the Main Noun was disambiguated (corresponding to point (2) in *Figure 1*), and the immediately adjacent Spill-over region to account for effects of the processing difficulty in the Coherence cue region "spilling over" into the next word. Eye-tracking allows for the measurement of natural reading behaviour without interrupting the reading process. An explicit Meaning Coherence Judgement task was used, in preference to silent reading interspersed with comprehension questions on filler trials. The choice of design meant that reading goals were controlled across all participants and stimuli, task difficulty remained relatively stable across trials, confounds introduced by task-switching were avoided, and the task allowed the collection of a direct outcome measure of sentence comprehension. To investigate the question of whether readers' lexical expertise affected on-line reading behaviour or off-line sentence comprehension for sentences that required reinterpretation, we also included measures of Vocabulary knowledge and Print Exposure. With these measures, we were able to capture two potentially separate aspects of lexical expertise: the depth of readers' word knowledge, and their experience with written material. By relating both measures to reader's eye movement behaviour and comprehension accuracy, we were able to explore potential links with knowledge and reading practice.

We hypothesised that processing costs related to reinterpretation would be observed. In line with the good-enough processing approach, we expected that readers might not always engage in reinterpretation, and therefore show lower accuracy on the Meaning Coherence Judgement task for Ambiguous compared to Unambiguous sentences (Ferreira, 2003; Ferreira & Patson, 2007). Based on previous evidence for ambiguity-related processing costs, we also expected to find longer reading times for Ambiguous compared to Unambiguous sentences (e.g. Duffy et al., 1988; Ferreira & Henderson, 1991; Rayner & Duffy, 1986; Rayner & Frazier, 1989; Rodd, Johnsrude, & Davis, 2010). In particular, it was expected that eye movements would reflect the detection of a processing difficulty at the Coherence cue region, and attempts at sentence reinterpretation. Apart from these group-level effects of Ambiguity on comprehension outcomes and on-line reading behaviour, we also hypothesised that we would find evidence for individual differences in reinterpretation costs. It was hypothesised that readers' lexical expertise would exert an influence both on their comprehension success in Ambiguous sentences and on their on-line reading behaviour during recovery from the misinterpretation. As we did not have specific hypotheses about the relative contributions of Vocabulary knowledge versus Print exposure, both measures were included in our statistical models. Based on previous findings, we expected to find evidence for facilitated lexical access in readers with greater lexical expertise (Ashby et al., 2005). The Lexical Quality Hypothesis additionally predicts an association of greater lexical expertise with facilitated incremental word-to-text integration processes, potentially leading to a benefit for the comprehension of sentences that require reinterpretation. Such a benefit would be promising for the development of potential interventions based on lexical expertise as a simple way to improve reader's ability to recover from misinterpretations.

Methods

Participants

Participants were recruited from the University of California, Davis subject pool. Data from 96 native speakers of American English (75 female, M_{Age} = 20 years, range: 18-34 years), who did not self-identify as bilingual and had normal or corrected-to-normal vision, were included in the analyses. An additional 5 participants were excluded from analyses because they were non-native speakers of English or had an uncorrected visual impariment, or due to procedural or technical difficulties with eye-tracking data collection. Written informed consent was obtained from all participants, and they were rewarded with university credits for participation. Ethical approval for this study was obtained from University of California, Davis, IRB 747381-11.

Materials and design

To investigate the recovery from misinterpretations during sentence comprehension and on-line processing, the critical sentence stimuli were designed to reliably lead readers down an interpretation "garden path". The first half of the sentence contained an ambiguous Main Noun with two meanings of unequal occurrence frequency (i.e. one dominant, relatively more frequent, and one subordinate, less frequent meaning, e.g. "ball")[1]. In the second half of the

---

[1] Although care was taken to choose ambiguous words with one clearly dominant meaning, other constraints for creation of our stimuli meant that some of our ambiguous Main Nouns had a more balanced meaning frequency

sentence, a single-word Coherence cue disambiguated the ambiguous Main Noun towards its unexpected, subordinate meaning (e.g. “dance event”). The Coherence cue region was therefore expected to coincide with the detection of a processing difficulty, and the initiation of reinterpretation processes. We will refer to this critical condition as the Ambiguous condition (see Example in *Table 1*). To compare comprehension and reading behaviour in these critical sentences to sentences that did not require reinterpretation, matched unambiguous sentence stimuli were created. The Main Noun was replaced with an unambiguous word whose meaning was compatible with the Coherence cue. Importantly, the remainder of the sentence content, including the Coherence cues, was unaltered in the Unambiguous condition. In all sentences, Main Noun and Coherence cue were separated by 4-8 words in order to ensure the completion of the meaning selection process (i.e. the selection of the dominant meaning) before the disambiguating Coherence cue was encountered (see Rodd, Johnsrude, & Davis, 2010; Seidenberg et al., 1982; Swinney, 1979). A neutral phrase was added to the end of each sentence to avoid contaminating the effects of the manipulated variables with processes associated with wrap-up effects (Warren, White, & Reichle, 2009), and programming of a return-sweep saccade.

---

distribution than others. However, group-level analyses of Ambiguity effects (see *Results*) indicated that overall, our stimulus selection was successful in generating garden-path effects.

Condition						
		Main Noun		***Coherence cue***	***Spill-over***	***Wrap-up***
Coherent Unambiguous	Sally worried that the	town	was going to be too	crowded	for her	liking.
Coherent Ambiguous	Sally worried that the	ball	was going to be too	crowded	for her	liking.
Anomalous Unambiguous	Sally worried that the	bird	was going to be too	crowded	for her	liking.
Anomalous Ambiguous	Sally worried that the	rule	was going to be too	crowded	for her	liking.

Table 1. **Example stimuli and sentence regions**. Coherent sentences either contained an ambiguous Main Noun (e.g. “ball”) that was disambiguated towards its subordinate meaning (“dance event”) by the Coherence cue region, or a matched unambiguous Main Noun (e.g. “town”), whose meaning was compatible with the Coherence cue. Semantically anomalous sentences contained either an ambiguous or unambiguous Main Noun which was not compatible with the meaning of the Coherence cue. The columns illustrate the sentence regions as defined for the eye-tracking analysis.

In order to avoid response bias in the Meaning Coherence Judgement task, additional sentences that were semantically anomalous, and therefore clearly required a “nonsense” response, were included. Anomalous sentences contained either an ambiguous or an unambiguous Main Noun to fully cross the factors of semantic Coherence (Coherent vs Anomalous) and Main Noun Ambiguity (Ambiguous vs Unambiguous). Main Nouns were chosen so that none of their meanings were semantically compatible with the Coherence cue. In summary, a quartet of matched sentence stimuli, which differed only in their Main Nouns, was created from each basic sentence to generate the four conditions (see *Table 1* for examples, and *Appendix I* for a full stimulus list). Main Nouns were not duplicated across conditions. The Main Nouns in the four conditions did not differ significantly in terms of their lemma frequency (based on the CELEX lexical database, Baayen, Piepenbrock, & Gulikers, 1995; $F(3, 188) = 0.65$, $p = .582$), their number of syllables ($F(3, 188) = 0.87$, $p = .457$), or number of letters ($F(3, 188) = 1.28$, $p = .284$; see descriptive statistics in *Table 2*).

The main purpose of this study was a comparison between sentences that required recovery from a misinterpretation (the Ambiguous condition), and those that did not require reinterpretation (the Unambiguous condition). Descriptive statistics for the Anomalous conditions are reported for completeness in the *Results* section, but are not analysed.

Condition	**Main Noun lemma frequency**	**Main Noun number of syllables**	**Main Noun number of letters**
Coherent Unambiguous	60.04 *(59.37)*	1.52 *(0.68)*	5.47 *(1.62)*
Coherent Ambiguous	74.42 *(75.88)*	1.56 *(1.62)*	4.96 *(1.38)*
Anomalous Unambiguous	56.65 *(48.68)*	1.33 *(0.48)*	5.02 *(1.51)*
Anomalous Ambiguous	68.31 *(85.82)*	1.31 *(0.51)*	5.06 *(1.26)*

Table 2. **Characteristics of stimuli used in the Meaning Coherence Judgement task**. Means (SD) of lemma frequency per million (CELEX lexical database, Baayen, Piepenbrock, & Gulikers, 1995), number of syllables, and number of letters of Main Nouns in each of the conditions in the Meaning Coherence Judgement task.

Coherent Ambiguous sentences came from a stimulus pool that has been used for previous experiments by our group (see e.g. Vitello, Warren, Devlin, & Rodd, 2014). They were selected based on the results of extensive piloting with British participants, which confirmed the relative meaning dominance of the ambiguous words both in isolation and in the particular sentence context, and ensured the comprehensibility of the sentences (see Supplementary Methods for details). To confirm the suitability of these stimuli for American English speakers, a two-stage pilot on single words and the sentence contexts was conducted specifically for the present experiment. In the first pilot stage, 33 participants from the University of California, Davis subject pool (who did not take part in the subsequent eye tracking experiment), were presented with an original set of 54 ambiguous words in isolation and ask to list all the meanings of the word they could think of. Responses were coded by LMR to indicate which meaning they referred to; coding was double-checked by research assistants who were naïve to the purpose of this experiment. This procedure allowed us to confirm whether American English speakers from the same subject pool as our target participants were familiar with both relevant meanings of the ambiguous nouns.

In the second pilot stage, 5 University of California, Davis research assistants who were native speakers of American English and naïve to the purpose of this experiment validated comprehensibility and naturalness of the sentence frames. On the basis of this two-stage piloting, 48 sentence quartets were selected, resulting in a total of 48 sentences per condition. These were divided into four lists, each containing 12 stimuli per condition; list presentation was counterbalanced across participants. Sentences from the same sentence frame quartet were not assigned to the same list, meaning that participants encountered each sentence frame in only one condition. Stimuli were presented in the same pseudorandomised order for each participant, following recommendations for individual differences designs (Swets, Desmet, Clifton, & Ferreira, 2008). Dependent variables were measured in three sentence regions of interest: the Main Noun, the Coherence Cue, and a Spill-over region, which directly followed the disambiguating information (see illustration in *Table 1*). Eye-tracking measures used in this study were defined as follows:

- First-fixation duration: The duration of the first fixation event that occurred within the area of interest (ms).
- Gaze duration: The summation of the duration of all fixations on the area of interest during first-pass reading (ms).
- Go-past time: The sum of the duration of all fixations on the area of interest from when it was fixated for the first time and until it was exited to the right for the first time (ms).
- Regressions out: A regression is a leftward eye movement leaving the area of interest to the left, before the area was first left to the right. For logistic mixed effects models, each trial was coded for the presence (1) or absence (0) of a regression. Summary measures across participants capture the total proportion of trials in a given condition in which one or multiple regressive eye movements occurred.
- Second-pass reading time: The sum of the duration of all fixations on the area of interest during second-pass reading of the sentence only (ms).

Apparatus

Eye movements were recorded with an SR Research EyeLink 1000 (SR Research Ltd, Ontario, Canada) desktop-mounted eye-tracking system with a sampling frequency of 1000 Hz. Sentences were presented on a 21 inch monitor (ViewSonic graphic series G225f) with 1024 x 768 pixels and a refresh rate of 75 Hz. Sentences were presented in black on a light grey background using 13 point Consolas font. Participants were seated 86 cm from the screen, with a chin rest to stabilise the head. Viewing was binocular but only one eye was tracked. For most participants, this was the right eye; due to instability in the data, one participant's left eye was tracked instead. The experiment was conducted using Experiment Builder software by SR Research.

Procedure

Participants were informed of the general purpose of the study (reading comprehension) and the nature of the Meaning Coherence Judgement task but were not informed of the ambiguity manipulation. The eye tracker was calibrated with a 9-point calibration procedure prior to the main experiment. Calibration was repeated during the experiment if necessary.

Participants completed the Meaning Coherence Judgement task on written sentences while their eye movements were recorded. At the beginning of each trial, a fixation point was displayed on the left-hand side of the screen, in approximately the location where the first word of the sentence would appear. Once the participant fixated on this point, the experimenter initiated the trial manually, and a sentence appeared on the screen. Participants were instructed to read silently at a normal pace for comprehension, and to decide as quickly and as accurately as possible whether each sentence made sense or did not make sense. These Meaning Coherence decisions were indicated using a button box. Participants were instructed to use their right index finger to press the left/green button for “Yes, this sentence makes sense”, and their middle or ring finger to press the right/red button for “No, this sentence does not make sense”. Most participants chose to respond using their right hand. One participant was left-handed and used the button box with their left hand; five further participants were left-handed but preferred to use their right hand for the task. Once the button was pressed, the sentence disappeared, and the next trial began. Before the experimental trials, participants practised the task on three unambiguous and coherent, and two unambiguous but anomalous sentences that were not repeated in the main task. Overall the duration of the Meaning Coherence Judgement was approximately 10 minutes.

The Meaning Coherence Judgement task was followed immediately by the Nelson-Denny vocabulary test, a multiple-choice measure of vocabulary knowledge. A digitised version of the task was presented on a computer screen, and participants marked their responses manually on a pre-printed response sheet. This task had a 15-minute time limit. Immediately after the vocabulary test, participants completed the Author Recognition task to probe print exposure (Acheson, Wells, & MacDonald, 2008; Stanovich & West, 1989). In this pen-and-paper task, participants were required to identify book authors among a list of names. Participants were instructed to avoid guessing, and to mark only those names they could confidently identify as book authors. This test has previously been shown to correlate with orthographic processing, vocabulary size, reading comprehension, and general knowledge (Stanovich & Cunningham, 1992; Stanovich & West, 1989), and is a measure that reduces social desirability effects as it asks participants to mark authors they recognise rather than asking for their reading behaviours more directly. There was no time limit for this task.

Data analysis

Within SR Research's DataViewer software, consecutive fixations that were shorter than 80 ms were merged, and single fixations shorter than 80 ms were removed prior to analysis. Such short fixations are unlikely to meaningfully contribute to cognitive processing (Rayner, 1998). As we were interested in inter-individual variability of reading behaviour, and we were concerned about biasing our data, reading times were not excluded for being too long (Ulrich & Miller, 1994). Eye-tracking data were manually checked for tracking quality by a research assistant who was naïve to the purpose and hypotheses of the study. No individual trials were excluded from the analyses. For all analyses, one item was removed ("bridge"). Despite the favourable piloting results for this item, its ambiguous sentence version (which relied on the relatively infrequent "game" meaning) was only judged to be meaningful by a single participant in the main experiment. Due to an error in the coding of Coherence cue and Spill-over regions-of-interest for the item "log", this item was also excluded from analyses of these regions.

Analyses were conducted in RStudio (v. 3.4.2, RStudio Team, 2015). Continuous eye-tracking measures were log-transformed for all analyses. Performance on the Nelson-Denny vocabulary test was scored as the total number of correct responses for each participant. For the Author Recognition Test, scores were calculated by subtracting false alarms (names incorrectly identified as authors) from hits (names correctly identified as authors) for each participant. Both the vocabulary and the Author Recognition scores were centred and scaled (i.e. standardised) prior to analysis.

Mixed effects models were fitted with random effects for subjects and items (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013). These models allow for simultaneous modelling of by-subject and by-item variance at the individual trial level – a clear advantage over the traditional approach of using separate by-subject and by-item ANOVAs. Following procedures outlined in Barr and colleagues (2013) for confirmatory hypothesis tests, maximal models were fitted where possible, and model complexity was reduced only where necessary. Parameter estimates from the maximal model, and Chi-squared values and *p*-values (α set at .05, or Bonferroni-corrected to .01 where indicated) from model comparisons with likelihood ratio tests are reported.

The main aim of this study was to investigate a) effects of Ambiguity (i.e. the requirement to recover from a misinterpretation) on on-line reading behaviour and comprehension outcomes of sentences that either made sense immediately (Coherent Unambiguous condition) or required reinterpretation processes (Coherent Ambiguous condition), and b) the role of individual differences in lexical expertise in the recovery from those misinterpretations. For this purpose, the analyses reported below were conducted on trials from the Coherent Ambiguous and Coherent Unambiguous conditions only, except where indicated otherwise. In the analyses of reading time measures, only correct trials were included to ensure that statistical inferences about reading behaviour were made only based on sentences for which participants were able to construct a coherent meaning representation. The models included fixed effects for Ambiguity (deviation-coded variables: Ambiguous -0.5, Unambiguous 0.5), for the continuous measures of Vocabulary and Print Exposure, and their interactions (Ambiguity x Vocabulary, and Ambiguity x Print Exposure). Random effects included a random intercept and random slopes for Vocabulary and Print Exposure by items, and a random intercept and random slope for Ambiguity by subjects.

The lmer() function of the lme4 package (v. 1.1.17; Bates et al., 2015) was used to fit linear mixed effects models for continuous dependent variables (first-fixation duration, gaze duration, go-past time, and second-pass reading time). For logit mixed effects models (accuracy of Meaning Coherence judgements, and probability of regressions out of an area of interest), the glmer() function within the same package was used with the bobyqa optimiser. Separate models were fitted for each region-of-interest and eye-tracking measure. As the analyses of eye tracking measures involved multiple comparisons, a Bonferroni correction was applied to these analyses with the α-level set to .01 (von der Malsburg & Angele, 2017). The data and code used for the analyses can be found on the OSF site associated with a preprint of this article (https://osf.io/hn3bu/).

Results

Reinterpretation demands in semantic garden-paths

Comprehension outcomes

Accuracy rates of Meaning Coherence judgements were significantly lower in the Coherent Ambiguous condition than in the Coherent Unambiguous condition (β = 3.5, SE = 0.26, z = 13.26; $\chi^2(1) = 110.62$, $p < .001$; see *Table 3* for descriptive statistics). On average, participants made 29% more errors in the Coherent Ambiguous condition than in the Coherent Unambiguous condition, indicating that readers did not always successfully recovery from their initial misinterpretation in the Coherent Ambiguous condition.

Total sentence reading time

Even when readers successfully comprehended the sentences, there was a significant difference in total reading times between Coherent Ambiguous and Coherent Unambiguous sentences (β = -0.05, SE = 0.004, t = -11.6; $\chi^2(1) = 85.04$, $p < .001$; see *Table 3* for descriptive statistics). Compared to processing of sentences that immediately made sense, successful recovery from an initial misinterpretation required additional processing time of about 440 ms on average.

Condition	**Accuracy (%)**	**Total reading time (ms)**
Coherent Unambiguous	92.2 *(26.8)*	3418 *(1253)*
Coherent Ambiguous	63.6 *(48.1)*	3859 *(1459)*
Anomalous Unambiguous	85.6 *(35.1)*	3815 *(1474)*
Anomalous Ambiguous	78.9 *(40.8)*	4075 *(1636)*

***Table 3.* Descriptive statistics for accuracy and total sentence reading time.** Means (*SD*) for accuracy rates, and for total sentence reading time (correct trials only). Although not included in our analyses, descriptive statistics for the Anomalous Unambiguous and Anomalous Ambiguous conditions are reported for completeness.

		Keyword	Coherence cue	Spill-over	Wrap-up
First-fixation duration (ms)					
	Unambiguous	235.4 *(85.7)*			
	Ambiguous	241.4 *(87.5)*			
	Coherent Unambiguous		241.4 *(80.6)*	271.6 *(121.9)*	256.9 (125.6)
	Coherent Ambiguous		241.2 *(79.0)*	266.6 *(108.7)*	259.7 (129.4)
	Anomalous Unambiguous		247.7 *(79.7)*	264.4 *(111.4)*	242.5 (112.9)
	Anomalous Ambiguous		245.1 *(74.9)*	261.0 *(124.6)*	249.4 (125.1)
Gaze duration (ms)					
	Unambiguous	257.4 *(115.0)*			
	Ambiguous	260.7 *(107.8)*			
	Coherent Unambiguous		275.6 *(126.9)*	362.4 *(224.5)*	322.1 (208.2)
	Coherent Ambiguous		276.1 *(135.5)*	370.4 *(206.5)*	338.7 (228.6)
	Anomalous Unambiguous		286.7 *(124.2)*	357.3 *(199.6)*	295.7 (182.5)
	Anomalous Ambiguous		285.5 *(135.6)*	348.8 *(211.8)*	311.2 (203.7)
Regressions out (probability)					
	Unambiguous	0.16 *(0.36)*			
	Ambiguous	0.14 *(0.35)*			
	Coherent Unambiguous		0.13 *(0.33)*	0.32 *(0.47)*	0.66 (0.47)
	Coherent Ambiguous		0.16 *(0.36)*	0.39 *(0.49)*	0.77 (0.42)
	Anomalous Unambiguous		0.13 *(0.34)*	0.44 *(0.5)*	0.77 (0.42)
	Anomalous Ambiguous		0.14 *(0.35)*	0.45 *(0.5)*	0.80 (0.4)
Go-past time (ms)					
	Unambiguous	328.6 *(250.0)*			
	Ambiguous	332.9 *(235.4)*			
	Coherent Unambiguous		335.5 *(261.0)*	631.4 *(644.3)*	823 (696.7)
	Coherent Ambiguous		354.0 *(262.4)*	792.7 *(822.0)*	1194.6 (900)
	Anomalous Unambiguous		358.2 *(287.7)*	770.0 *(765.7)*	1067.9 (878)
	Anomalous Ambiguous		380.2 *(350.9)*	866.0 *(938.2)*	1223.4 (1030)

Second-pass reading time (ms)					
	Coherent Unambiguous	246.8 *(128.3)*	252.7 *(174.3)*	310.6 *(185.7)*	326 (249.8)
	Coherent Ambiguous	280.6 *(150.2)*	271.4 *(153.8)*	296.9 *(186.8)*	315.9 (211.2)
	Anomalous Unambiguous	266.3 *(127.2)*	266.1 *(140.2)*	306.5 *(204.8)*	276.2 (163)
	Anomalous Ambiguous	278.2 *(148.4)*	289.0 *(165.6)*	293.7 *(180.1)*	294.8 (176.5)

Table 4. **Descriptive statistics for eye tracking measures.** Means (*SD*) for correct trials are reported for the combined unambiguous conditions and combined ambiguous conditions for first-pass measures in the Main Noun region, and for all four conditions separately in other sentence regions. Descriptive statistics for the Anomalous Ambiguous and Anomalous Unambiguous conditions are reported for completeness.

On-line processing

First encounter of an ambiguous vs unambiguous word

Because Ambiguous and Unambiguous sentences were not differentiated in terms of meaning coherence until the Coherence cue region, first-pass measures in the Main Noun region were combined across both Ambiguous conditions (Coherent Ambiguous and Anomalous Ambiguous) and both Unambiguous conditions (Coherent Unambiguous and Anomalous Unambiguous) for comparison, which maximised statistical power.[2] Effects of Ambiguity on first-fixation duration, gaze duration, the probability of making a regression out of the Main Noun region into earlier parts of the sentence, or on go-past time either did not survive correction for multiple comparisons, or were non-significant (see *Table 4* for descriptive statistics and *Appendix II Table 2*). These results thus suggest that upon first encounter of the Main Noun region, reading behaviour was not affected by whether or not the word had a single or multiple meanings.

---

[2] Main Noun region analyses conducted on only the Coherent Ambiguous and Coherent Unambiguous trials demonstrated statistically weaker but overall similar results.

Reaction to a processing difficulty

Reading behaviour indicative of processing difficulty was expected to occur in the Coherence cue region, which contained information that disambiguated the ambiguous Main Noun towards an unexpected meaning. The Spill-over region was analysed to detect any effects triggered by the disambiguating information that carried over to the immediately adjacent sentence region. There were no significant effects of Ambiguity in the Coherence cue region (see *Appendix II Table 4*). However, in the Spill-over region, there was a significant effect of Ambiguity on the probability of making a regressive eye movement out of the region towards earlier parts of the sentence (β = -0.42, SE = 0.13, z = -3.32; $\chi^2(1) = 10.63$, $p < .001$). Readers tended to make more regressive eye movements out of the Spill-over region in the Coherent Ambiguous compared to the Coherent Unambiguous condition. Ambiguity also had a significant effect on go-past time in the Spill-over region, with readers taking longer to move out of the Spill-over to subsequent sentence regions when the sentence was ambiguous (β = -0.07, SE = 0.01, t = -4.86; $\chi^2(1) = 23.04$, $p < .001$). Evidence for processing difficulty in the semantic garden-path sentences was therefore found not immediately upon encounter of the disambiguating information, but in the adjacent sentence region. Processing difficulty was reflected in the eyes lingering on the problematic area, and the initiation of saccades towards earlier parts of the sentence.

To further explore the effects of Ambiguity on reading behaviour after encountering a processing difficulty, secondary analyses were conducted on Ambiguous and Unambiguous conditions in the Wrap-up region. There was a significant effect of Ambiguity on go-past time in the Wrap-up region (β = -0.18, SE = 0.02, t = -9.89; $\chi^2(1) = 62.88$, $p < .001$), suggesting that readers spent longer on the final portion of the sentence when the sentence required reinterpretation. Similarly, the probability of making a regression out of the Wrap-up region was significantly higher in the Ambiguous than in Unambiguous sentences, echoing the results from the Spill-over region (β = -0.77, SE = 0.18, z = -4.21; $\chi^2(1) = 17.81$, $p < .001$, see *Appendix II Table 8*).

Re-reading behaviour

In addition to the first-pass through the sentence, we also analysed second-pass reading times in our sentence regions-of-interest. Results indicated that readers revisited the Main Noun region for longer in the Coherent Ambiguous than in the Coherent Unambiguous condition by on average around 34 ms (see *Table 4* for descriptive statistics; β = -0.05, SE = 0.02, t = -2.72; $\chi^2(1) = 7.22$, $p = .007$). A similar effect of Ambiguity on second-pass reading time in the Coherence cue region did not survive correction for multiple comparisons (β = -0.05, SE = 0.02, t = -2.2; $\chi^2(1) = 4.79$, $p = .029$), and the effect of Ambiguity on second-pass reading time in the Spill-over region was non-significant (see *Appendix II Table 5*).

Individual differences in lexical expertise

Participants' scores on the Nelson-Denny test (Vocabulary knowledge) ranged between 41 and 78, out of a maximum score of 80 ($M_{Vocab} = 64.4$, $SD_{Vocab} = 7.7$). Although the mean score on the Author Recognition test was quite low in our sample ($M_{Print} = 13.0$, $SD_{Print} = 7.4$, range = 0 – 41, out of a maximum score of 64), scores showed an approximately normal distribution across the lower third of the range of possible scores (see *Figure 2*). Two participants seemed to score unusually high on this task compared to the rest of our sample; however, since these scores were at ceiling but were not based on consistent selection of the same response option, we believe that they are a genuine reflection of participants' print exposure. We therefore did not remove these scores from our analysis. Performance on the Vocabulary and Print Exposure measures was moderately correlated, $r = .53$, $p < .001$. However, as there was no indication of concerning levels of collinearity between the two variables (Variance Inflation Factors <4, see O'Brien, 2007, see App. II Table 9), both were included in our statistical models.

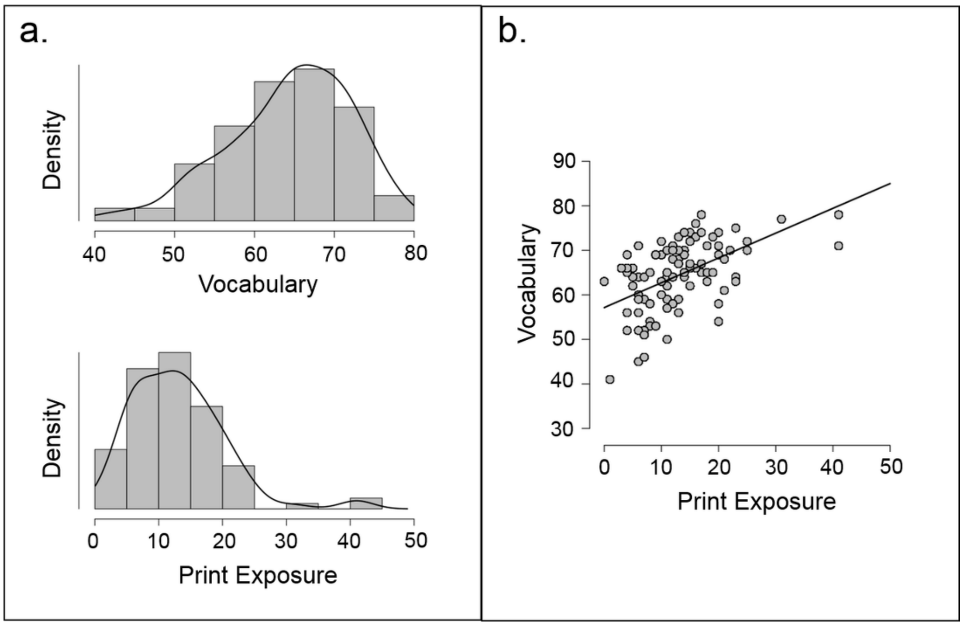

Figure 2. Relationship between participants' Vocabulary and Print Exposure scores. Panel a. shows distribution plots for performance on the Vocabulary and Print Exposure measures. Panel b. shows the correlation (r = 0.53, p < .001) between the two measures.

To be able to draw meaningful conclusions from a correlational investigation of individual differences, our dependent variables should show variability in Ambiguity effects across individuals (Spearman, 1904). As can be seen in *Figure 3*, the difference between mean eye-tracking measures for Ambiguous compared to Unambiguous trials varied considerably across individuals. As the main goal of this experiment was to investigate the role of lexical expertise specifically for the recovery from misinterpretations, we will discuss effects of the lexical expertise variables and their interactions with Ambiguity only for measures where a group-level effect of Ambiguity has been observed. Results for all measures can be found in the tables in the Appendix.

Comprehension outcomes

Effects of Vocabulary, Print Exposure, and their interactions with Ambiguity on accuracy rates on the Meaning Coherence Judgement task were non-significant (see *Appendix II Table 1*).The outcome of comprehension processes was therefore not affected by readers' lexical expertise. Importantly, the comprehension detriment due reinterpretation demands in the

Coherent Ambiguous sentences was relatively consistent across our sample, and did not depend on an individual’s Vocabulary or Print Exposure scores.

Total sentence reading time

Although higher Print Exposure scores were associated with faster total sentence reading time (β = -0.03, SE = 0.01, t = -2.3; $\chi^2(1)$ = 5.27, *p* = .022), interactions of Ambiguity x Vocabulary and Ambiguity x Print Exposure were non-significant (see *Appendix II Table 1*). The cost of reinterpretation demands to reading time in the Coherent Ambiguous sentences did therefore not depend on readers’ lexical expertise.

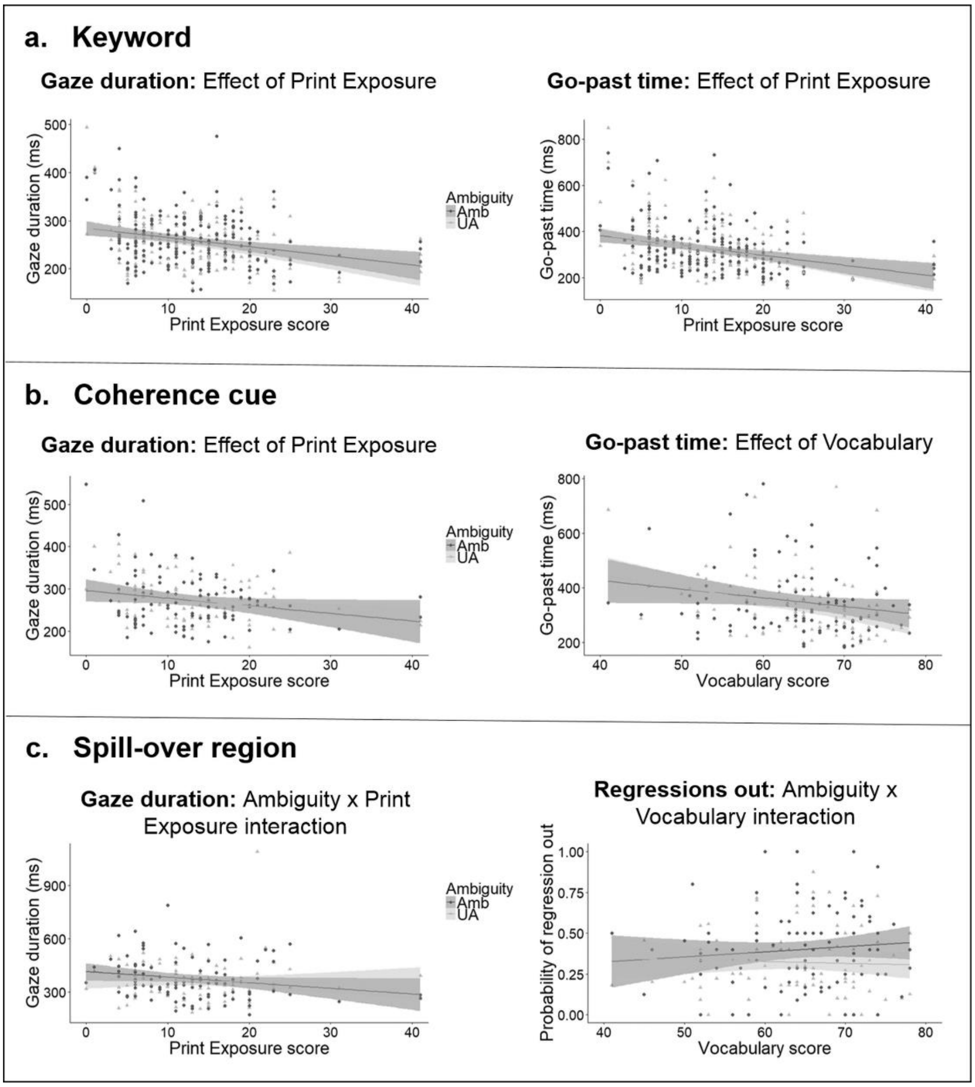

Figure 3. Relationship between lexical expertise and eye tracking measures in different sentence regions. Scatterplots for gaze duration, and go-past time in the Main Noun region (where no differences between conditions were expected) show a comparison between combined ambiguous conditions (Coherent Ambiguous and

Anomalous Ambiguous) and combined unambiguous conditions (Coherent Unambiguous and Anomalous Unambiguous), reflecting the analyses performed (panel a.). Scatterplots for all other measures compare the Coherent Ambiguous to the Coherent Unambiguous condition (panels b. and c.).

On-line processing

Reaction to a processing difficulty

Analyses of the Spill-over region showed some indication that the observed main effect of Ambiguity on reading behaviour was in fact modulated by readers' lexical expertise (see *Figure 3*, and *Appendix II Table 5*). Although the Ambiguity x Vocabulary interaction on the probability of making a regression out of the Spill-over region did not reach a corrected significance level (β = -0.3, SE = 0.15, z = -2.06; $\chi^2(1) = 4.15$, p = .042), the results from pairwise comparisons are suggestive of a greater divergence between the likelihood of making a regression out of the Spill-over region of a Coherent Ambiguous compared to a Coherent Unambiguous sentence with greater Vocabulary scores (see *Figure 3*, and *Appendix II Table 6*).

To further explore the effects of lexical expertise and Ambiguity on reading behaviour after encountering a processing difficulty, secondary analyses were conducted on the Wrap-up region. There was a significant effect of Print Exposure on gaze durations (β = -0.04, SE = 0.02, t = -2.81; $\chi^2(1) = 7.28$, p = .007), suggesting that readers with higher Print Exposure scores tended to have shorter gaze durations in this region. However, the effects of reinterpretation demands on reading behaviour in the Wrap-up region were not modulated by readers' lexical expertise (see *Appendix II Table 8).*

Re-reading behaviour

Effects of Ambiguity on second-pass reading time in all regions were relatively stable across the sample, and were not affected by readers' lexical expertise (see *Appendix II Tables 2, 4, and 5*).

Discussion

The present study investigated both the outcome of comprehension procedures, and on-line reading behaviour during processing of sentences that require the recovery from misinterpretations. We hypothesised that semantic garden-path sentences would be associated with lower comprehension accuracy compared to matched unambiguous sentences, and that the detection of processing errors and related recovery procedures would come with reading time costs and eye movement behaviour consistent with reinterpretation attempts. In addition, we examined whether individual differences in lexical expertise affected comprehension outcomes and reinterpretation procedures during online processing.

In line with our hypotheses, results from the present experiment revealed robust group-level effects of Ambiguity. Accuracy rates on the Meaning Coherence Judgement task were lower for Ambiguous compared to Unambiguous sentences. Participants did not always resolve the ambiguities in the semantic garden-path sentences, and therefore erroneously declared them nonsensical on 37% of trials on average. This finding is in line with previous work suggesting that even skilled readers do not always resolve ambiguities, and therefore consistent with the idea that comprehenders engage in processing strategies that are just “good enough” to fulfil the task at hand (Christianson et al., 2001; Ferreira, 2003; Ferreira & Patson, 2007). An alternative explanation for the comparatively low accuracy rates on the Meaning Coherence Judgement task for Ambiguous sentences is that readers were unable to resolve the ambiguities in some sentences because they may not have been familiar with the subordinate meaning of the word. While infrequent in comparison to the dominant meanings, the subordinate meanings used in our stimuli were still commonly used. Pilot testing of the stimuli also indicated that individuals from the same participant pool were familiar with both meanings of the ambiguous words. Apart from one item, which was subsequently removed from the analyses (*bridge*), none of our sentences were consistently misinterpreted by the majority of our participants. We therefore believe that the present data are better interpreted as a case of “good enough” processing. It is likely that the use of a good enough processing strategy was encouraged in this task as participants also encountered semantically anomalous sentences where reinterpretation attempts were futile. Although it is difficult to develop tasks that provide a measure of comprehension outcomes but do not affect the type of processing in which readers engage, future research may want to investigate the effects of the Meaning Coherence Judgement task on processing strategies (see Swets et al., 2008, for effects of task difficulty on the use of good enough processing).

Evidence from total sentence reading time and the on-line eye-tracking measures further indicated that successful recovery from misinterpretations was associated with processing costs. Even on trials that were correctly judged as coherent, readers took significantly longer to arrive at the correct interpretation of Ambiguous compared to Unambiguous sentences by 440 ms on average, as indicated by the analyses of total sentence reading time. This result suggests that the detection of processing errors and completion of reinterpretation procedures required additional processing time, replicating previous findings of temporal processing costs associated with successful ambiguity resolution (Duffy et al., 1988; Kambe, Rayner, & Duffy, 2001; Rodd, Johnsrude, & Davis, 2010). On-line eye movement measures further indicated that sentences containing ambiguous nouns led our participants down an interpretation garden path. At the Main Noun region, reading measures for Ambiguous and Unambiguous words did not differ. This finding was expected as the stimuli used in this study contained ambiguous words with unequal distributions of relative meaning frequency, and provided disambiguating context for the ambiguous word only later in the sentence. Consistent with theories that view ambiguity resolution as the result of an interaction of contextual cues and the relative frequency of the meanings of an ambiguous word (e.g. Duffy et al., 1988; Simpson, 1981; Twilley & Dixon, 2000), it was therefore predicted that initially, the (incorrect) dominant meaning of the ambiguous word would be selected based on its greater relative occurrence frequency. Similar to the present study, previous research has found that dominant meanings of ambiguous words are accessed at a similar speed as unambiguous word meanings, indicating that meaning *selection* occurs very rapidly and without the need for additional processing time over and above the time needed for meaning *access* (Duffy et al., 1988; Rodd et al., 2010).

The sentences used in this study were constructed such that readers would detect a violation to meaning coherence further downstream from the ambiguous word, upon encounter of the disambiguating information in the Coherence cue region. In line with previous research on processing difficulties during reading, we predicted that at the point of disambiguation, the detection of a coherence violation would be indexed by a difference in first-pass reading measures (e.g. first-fixation duration, gaze duration, or go-past time) between the Ambiguous and the Unambiguous condition (Christianson, Luke, Hussey, & Wochna, 2017; Rayner, Warren, Juhasz, & Liversedge, 2004). However, eye movement measures in this region did not reveal Ambiguity effects consistent with the detection of a processing difficulty. Instead, such Ambiguity effects were found only in the subsequent Spill-over region immediately following the disambiguating information. One reason why readers in this study may not have detected a coherence violation in the Coherence cue region itself may be that the region consisted only of a single, short word (usually between one and two syllables); the cue for a coherence violation may therefore have been relatively weak at this stage of processing. Alternatively, the finding of significant Ambiguity effects in the Spill-over, rather than the Coherence cue region itself, may have been a consequence of the rate of saccade programming. The planning of a saccade out of the Coherence cue region may have taken less time than integration of the region into the preceding context (or rather, the failure of such integration processes in the Ambiguous condition). Indeed, the average gaze duration in the Coherence cue region (about 275 ms) was shorter than gaze duration in the Spill-over region (about 366 ms), suggesting that full processing of the content of the Coherence cue region happened after the programming of a saccade into the subsequent region. In the Spill-over region, go-past times (i.e. the total time spent in this region once it was entered from the left and before it was exited for the first time) were significantly longer in the Ambiguous compared to the Unambiguous condition, suggesting that readers lingered in this region when they had detected an error in their initial sentence interpretation. Additionally, there was a significantly greater proportion of trials (by about 7%) in which readers initiated a regression out of the Spill-over region in the Ambiguous compared to the Unambiguous condition. Ambiguity effects indicative of rereading were also found in the final portion of the sentence. Go-past times and regressions out of the Wrap-up region suggested that readers spent longer in the region and were more likely to move back towards earlier parts of the sentence when they had been led down a garden path. These findings suggest that readers reacted to having been led down a garden-path by making backwards eye movements returning to previous parts in the sentence in order to resolve the processing error. Reinterpretation processes in the present study were therefore associated with physical re-reading of the sentence. Analyses of second-pass reading times (i.e. time spent reading a sentence region for the second time) indicated that it was in particular the Main Noun region (rather than the Coherence cue or Spill-over regions) that was revisited when a misinterpretation had occurred. Although this finding is consistent with the idea that readers may have selectively focused their rereading efforts on the problematic ambiguous word to access and integrate an initially discarded meaning, it is important to note that we did not investigate rereading on the surrounding sentence regions (i.e. the initial sentence region, or the region directly following on from the ambiguous Main Noun). It is therefore an open question whether semantic garden-paths encourage a selective re-reading strategy (e.g. Frazier & Rayner, 1982), or a strategy whereby the whole sentence is reread and reinterpreted from scratch (see Lewis, 1998, for a discussion). It is likely that the nature of the decision that participants needed to make (the Meaning Coherence judgement), and the presence of semantically anomalous sentences in a portion of trials encouraged re-reading as a strategy to resolve semantic inconsistencies, and to “double-check” each decision. Even in the coherent Unambiguous sentences, participants were likely to engage in some re-reading, suggesting that checking sentence meanings was a general, possibly task-related, processing strategy. The Meaning Coherence judgement task may have overall encouraged physical rereading across the different conditions; however, the focus on meaningfulness in our task ensured motivation in our participants to resolve the ambiguities in the majority of trials, and therefore allowed us to investigate on-line reading behaviour associated with successful recovery from a misinterpretation (cf. Ferreira & Patson, 2007). A question to consider in future research is whether the present findings can be replicated under more natural reading conditions without the presence of an artificial task. In particular, a future study may want to investigate how the recovery from misinterpretations is affected by the availability of previous sentence content for rereading, for example by using a moving-window paradigm where prior sentence regions do not remain visible.

Processing costs associated with the recovery from misinterpretations were observed in a relatively large sample of adult readers. In addition to the group-level effects of Ambiguity, however, we also investigated the effects of individual readers' lexical expertise on their reading behaviour and comprehension accuracy. Although both vocabulary knowledge and print exposure were highly variable in the present sample of university-educated young adults, lexical expertise was not found to exert any effects on the accuracy of Meaning Coherence judgements. Perfetti & Stafura's (2014) Reading Systems Framework assumes that readers with weaker lexical representations (or lower levels of lexical expertise in the parlance of the present study) have to expend additional processing resources on comparatively simple procedures like lexical access and word-to-text integration. As processing resources are depleted by lexical access and integration, readers with lower lexical expertise are predicted to have fewer resources to spare for other procedures. In the present study, however, the consequences of resource depletion were not evident for off-line comprehension. As successful comprehension of temporarily ambiguous sentences required the detection of a coherence violation and its resolution by accessing and integrating a less frequent meaning of the ambiguous word, this finding suggests that readers with limited lexical expertise were generally able to complete those processes just as well as readers with greater lexical expertise. This finding was therefore in contrast to our hypothesis about the benefits of vocabulary knowledge and reading experience to the outcome of processes related to the recovery from misinterpretations. In contrast to the present results, previous research on younger readers has found that vocabulary size significantly affected the accuracy with which difficult sentences were understood (Engelhardt, Nigg, & Ferreira, 2017). However, adult readers may, as a function of their relatively greater reading experience compared to beginning readers, have developed strategies to compensate for limitations in their lexical expertise by modulating their reading or response speed to ensure good levels of accuracy; the significant relationship between lower lexical expertise and longer total sentence reading times in the present study is consistent with this explanation. Furthermore, Engelhardt and colleagues' (2017) measure of comprehension ability was the accuracy of responses to comprehension questions; responding to such questions required the *reconstruction* of sentence meaning, and may therefore have introduced additional skill requirements that were not measured in the present study. In sum, the present results suggest that the ultimate success of ambiguity resolution processes was not significantly influenced by adult readers' lexical expertise. Irrespective of their lexical expertise, readers in our sample tended to use a good-enough processing strategy, and did not always engage in reinterpretation processes to resolve the ambiguity in the semantic garden-paths.

When correct decisions about the sentences were made, total sentence reading time tended to be faster for individuals with greater Print exposure scores. Similarly, main effects of lexical expertise variables on eye-movement measures generally suggested a trend for an association of greater lexical expertise with more efficient lexical access and word-to-text integration processes (see Ashby et al., 2005 and Kuperman & Van Dyke, 2011 for similar results). These findings are in line with the predictions of the Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2002). However, benefits to reading time were observed irrespective of the presence of ambiguity in the sentences, suggesting that the costs of Ambiguity to processing time due to the detection of a processing error and its resolution were relatively stable across our sample.

Of key interest to the present purposes were interactions between Ambiguity and the lexical expertise measures. Before discussing the relevant results, it is important to note that we applied a relatively stringent significance threshold to our eye-tracking analyses following recommendations by von der Malsburg and Angele (2017). A Bonferroni-correction was applied to guard against inflated Type I errors due to multiple comparisons (i.e. the use of 5 related measures), so results that only met an uncorrected significance threshold should be interpreted with caution. The present study did not find evidence to suggest that readers with lower lexical expertise were *unable* to detect, or even resolve, coherence violations due to a misinterpretation. Instead, there was some evidence to suggest that greater lexical expertise may have specific benefits for the *efficiency* with which readers detected and resolved coherence violations caused by lexical ambiguity. In the Spill-over region, there was a trend towards an Ambiguity x Vocabulary knowledge interaction, suggesting that readers with greater lexical expertise engaged more selectively in overt re-reading of previous sentence content (i.e. greater probability of making a regression out of the Spill-over region) in the Ambiguous rather than Unambiguous sentences compared to readers with lower levels of lexical expertise. Neither reading behaviour in the final sentence region nor rereading of the Main Noun region were affected by lexical expertise. Overall, the present findings therefore suggest that readers with lower levels of vocabulary knowledge or reading experience can complete reinterpretation processes with as much success as “lexical experts”, and that they engage in similar reading strategies to recover from an initial misinterpretation. Crucially, however, at the Spill-over region, readers with greater lexical expertise experienced an exaggerated garden-path effect.

One explanation for the exaggerated garden-path effect in readers with greater lexical expertise is that their rapid word-to-text integration ability provides them with an advantage during regular sentence processing, but that their ability to build elaborated sentence-meaning representations leads to a disadvantage for the recovery from a misinterpretation. Since their initial interpretation is highly elaborated and well-integrated, a processing difficulty such as that introduced in our sentences may be harder to recover from for them compared to readers who do not generate similarly detailed sentence-level representations during processing (see discussion of “digging-in” effects in e.g. Tabor & Hutchins, 2004). However, the present results indicate that it was only the probability of making a regression out of the Spill-over region which distinguished the on-line reading behaviour of individuals with different levels of lexical expertise. In fact, regressions out of the following Wrap-up region were comparable for different readers, suggesting that readers with greater lexical expertise were able to engage in reinterpretation at an earlier point in the sentence compared to readers with lower levels of lexical expertise. In summary, processes related to the detection of coherence violations and re-reading behaviour *per se* were not affected by lexical expertise. The critical difference between readers of different levels of lexical expertise seemed to lie in the sensitivity to the presence of ambiguity upon encounter of the disambiguating information.

Similar to the present study, in a comparison of self-paced reading times for disambiguating regions in syntactically ambiguous sentences to matched regions in unambiguous control sentences, Pearlmutter and MacDonald (1995) found that high-span readers showed longer reading times for ambiguous than unambiguous sentences, while low-spans showed a smaller effect of ambiguity. The authors argued that differences between span groups were not due to differences in the ability to maintain multiple interpretations, but due to differences in sensitivity to probabilistic constraints offered by sentence contexts. In particular, they proposed that high-span readers may have more computational capacity available to detect subtle contextual constraints during online processing (see also Just & Carpenter, 1992, for a similar argument regarding inter-individual differences in sensitivity to constraints of agent animacy). Although working memory capacity was not measured in the present study, our results are remarkably similar to those reported by Pearlmutter and MacDonald (1995): readers with greater lexical expertise showed *greater* effects of ambiguity in the Spill-over region than readers with lower lexical expertise. In line with our results, Pearlmutter & MacDonald (1995) found no differences between span groups on offline comprehension measures; differences in sensitivity to contextual and probabilistic constraints only emerged in measures of online processing. Similarly, Long and Prat (2008) also found that high-span readers tended to be more sensitive to ambiguity at an earlier point during processing compared to low-spans. However, high- and low-span groups showed no differences in offline measures of comprehension. These findings suggest that although lower-skill readers (whether due to smaller reading span, limited lexical knowledge or lower levels of print exposure) can eventually arrive at a correct interpretation, they tend to be less sensitive to constraints offered by the sentence context during online sentence processing. An important avenue for future studies to explore is the relative contribution of print exposure versus vocabulary measures to the recovery from ambiguity-related misinterpretations individually, and potential common factors that underlie the effects of lexical expertise and working memory on online sentence processing (see e.g. Long & Prat, 2008; MacDonald & Christiansen, 2002; Wells et al., 2009).

Future research is needed to better understand the contribution of lexical knowledge and reading experience to reinterpretation processes. Although the present study suggests that lexical expertise did not influence our offline comprehension measure, we found indications that greater vocabulary knowledge may be associated with more efficient detection and resolution of processing difficulties during on-line processing. Research on readers with a wider range of lexical expertise, including participants from non-university backgrounds, is particularly important considering that the range of Print exposure scores in the present sample was limited (see e.g. Mainz et al., 2017). Vocabulary training or interventions based on reading practice are relatively simple to administer; developing a better understanding of the benefits of lexical expertise for comprehension and on-line processing of difficult linguistic material is therefore a promising avenue for future research.

Conclusions

The present study investigated comprehension outcomes and on-line processing during the recovery from misinterpretations, and the influence of individual differences in lexical expertise on reinterpretation processes in adult readers. Sentences that contained a temporary lexical-semantic ambiguity were not always correctly interpreted, and led readers down an interpretation garden-path that required additional processing time compared to matched unambiguous sentences. Although comprehension success and rereading strategies seemed to be comparable for readers with different levels of vocabulary knowledge and reading experience, there was some evidence to suggest that readers with greater lexical expertise may be more sensitive to the presence of ambiguity during processing, and therefore able to flexibly adapt their on-line reading behaviour to recover from misinterpretations.

Acknowledgements

This work was part of Lena Maria Blott's doctoral dissertation at University College London, which was supported by an ESRC-MRC Multidisciplinary Studentship. Data collection at University of California, Davis, was supported by a Bogue Fellowship to Lena Maria Blott, and by NIA Grant 1R56AG053346-01A1 to Fernanda Ferreira. The authors would like to thank Wallace Lau, John Hansen and Pierre Llorach for help with data collection, and Suphasiree Chantavarin, Gwen Brekelmans, and Eva D. Poort for help with Experiment Builder and stubborn code.

References

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, *40*(1), 278–289. https://doi.org/10.3758/BRM.40.1.278

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the Simple View of Reading Include a Fluency Component? *Reading and Writing*, *19*(9), 933–958. https://doi.org/10.1007/s11145-006-9024-z

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye Movements of Highly Skilled and Average Readers: Differential Effects of Frequency and Predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065–1086. https://doi.org/10.1080/02724980443000476

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/ 10.18637/jss.v067.i01

Blott, L. M., Rodd, J. M., Ferreira, F., & Warren, J. (2019, April 24). Recovery from misinterpretations during online sentence processing. https://doi.org/10.17605/OSF.IO/HN3BU

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking Up for Vocabulary: Reading Skill Differences in Young Adults. *Journal of Learning Disabilities*, *40*(3), 226–243. https://doi.org/10.1177/00222194070400030401

Cairns, H. S., & Kamerman, J. (1975). Lexical information processing during sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *14*(2), 170–179. https://doi.org/10.1016/S0022-5371(75)80063-6

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology, 42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(4), 561-584. http://dx.doi.org/10.1037/0278-7393.9.4.561

Dopkins, S., Morris, R. K., & Rayner, K. (1992). Lexical ambiguity and eye fixations in reading: A test of competing models of lexical ambiguity resolution. *Journal of Memory and Language*, *31*(4), 461–476. https://doi.org/10.1016/0749-596X(92)90023-Q

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*(4), 429–446. https://doi.org/10.1016/0749-596X(88)90066-6

Engelhardt, P. E., Nigg, J. T., & Ferreira, F. (2017). Executive Function and Intelligence in the Resolution of Temporary Syntactic Ambiguity: An Individual Differences Investigation.

Quarterly Journal of Experimental Psychology, *70*(7), 1263–1281. https://doi.org/10.1080/17470218.2016.1178785

Farmer, T. A., Misyak, J. B., & Christiansen, M. H. (2012). Individual Differences in Sentence Processing. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics* (1st ed., pp. 353–364). https://doi.org/10.1017/CBO9781139029377.018

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146. https://doi.org/10.1017/S0048577202001373

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Ferreira, F., & Clifton, C. (1986). The Independence of Syntactic Processing. *Journal of Memory and Language, 25*(3), 348-368. https://doi.org/10.1016/0749-596X(86)90006-9

Ferreira, F., & Henderson, J. M. (1991). Recovery from Misanalyses of Garden-Path Sentences. *Journal of Memory and Language, 30*(6), 725-745. https://doi.org/10.1016/0749-596X(91)90034-H

Ferreira, F., & Patson, N. D. (2007). The Good Enough Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Foss, D. J., Bever, T. G., & Silver, M. (1968). The comprehension and verification of ambiguous sentences. *Perception & Psychophysics*, *4*(5), 304–306. https://doi.org/10.3758/BF03210520

Foss, D. J., & Jenkins, C. M. (1973). Some effects of context on the comprehension of ambiguous sentences. *Journal of Verbal Learning and Verbal Behavior*, *12*(5), 577–589. https://doi.org/10.1016/S0022-5371(73)80037-4

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, *26*(5), 505–526. https://doi.org/10.1016/0749-596X(87)90137-9

Gadsby, N., Arnott, W. L., & Copland, D. A. (2008). An investigation of working memory influences on lexical ambiguity resolution. *Neuropsychology*, *22*(2), 209–216. https://doi.org/10.1037/0894-4105.22.2.209

Gernsbacher, M. A., & Faust, M. E. (1991). The Mechanism of Suppression: A Component of General Comprehension Skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(2), 245-262. http://dx.doi.org/10.1037/0278-7393.17.2.245

Gorfein, D.S., Viviani, J.M., & Leddo, J. (1982). Norms as a tool for the study of homography. *Memory & Cognition, 10*, 503-509.

Guo, Y., Roehrig, A. D., & Williams, R. S. (2011). The Relation of Morphological Awareness and Syntactic Awareness to Adults' Reading Comprehension: Is Vocabulary Knowledge a Mediating Variable? *Journal of Literacy Research*, *43*(2), 159-183. https://doi.org/10.1177/1086296X11403086

Hogaboam, T. W., & Perfetti, C. A. (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *14*(3), 265–274. https://doi.org/10.1016/S0022-5371(75)80070-3

Holmes, V. M. (1979). Accessing Ambiguous Words during Sentence Comprehension. *Quarterly Journal of Experimental Psychology*, *31*(4), 569–589. https://doi.org/10.1080/14640747908400749

Holmes, V. M., Arwas, R., & Garrett, M. F. (1977). Prior context and the perception of lexically ambiguous sentences. *Memory & Cognition*, *5*(1), 103–110. https://doi.org/10.3758/BF03209200

Joshi, R. M. (2005). Vocabulary: A Critical Component of Comprehension. *Reading & Writing Quarterly, 21*(3), 209–219. https://doi.org/10.1080/10573560590949278

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329-354. http://dx.doi.org/10.1037/0033-295X.87.4.329

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149. http://dx.doi.org/10.1037/0033-295X.99.1.122

Kambe, G., Rayner, K., & Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words: Evidence from eye fixations. *Memory & Cognition*, *29*(2), 363–372. https://doi.org/10.3758/BF03194931

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology, 69*(5), 1013–1040. https://doi.org/10.1080/17470218.2015.1053951

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*(1), 42–73. https://doi.org/10.1016/j.jml.2011.03.002

Lewis, R. L. (1998). Reanalysis and Limited Repair Parsing: Leaping off the Garden Path. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in Sentence Processing* (Vol. 21, pp. 247–285). https://doi.org/10.1007/978-94-015-9070-9_8

Long, D. L., & Prat, C. S. (2008). Individual differences in syntactic ambiguity resolution: Readers vary in their use of plausibility information. *Memory & Cognition, 36*(2), 375–391. https://doi.org/10.3758/MC.36.2.375

Luke, S. G., Henderson, J. M., & Ferreira, F. (2015). Children’s eye-movements during reading reflect the quality of lexical representations: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(6), 1675–1683. https://doi.org/10.1037/xlm0000133

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35–54. https://doi.org/10.1037//0033-295X.109.1.35

MacDonald, M. C., Pearlmutter, & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review, 101*(4), 676-703. http://dx.doi.org/10.1037/0033-295X.101.4.676

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary Knowledge Predicts Lexical Processing: Evidence from a Group of Participants with Diverse Educational Backgrounds. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01164

Nelson, D.L., MsEvoy, C.L., Walling, J.R., & Wheeler, J.W. (1980). The University of South Florida homograph norms. *Behavior Research Methods & Instrumentation, 12,* 16-37.

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, *9*(3), 225–236. https://doi.org/10.3758/BF03196957

Payne, B. R., Gao, X., Noh, S. R., Anderson, C. J., & Stine-Morrow, E. A. L. (2012). The effects of print exposure on sentence processing and memory in older adults: Evidence for efficiency and reserve. *Aging, Neuropsychology, and Cognition*, *19*(1–2), 122–149. https://doi.org/10.1080/13825585.2011.628376

Pearlmutter, & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language 34*(4), 521-542.

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Perfetti, C., & Hart, L. (2002). *The Lexical Quality Hypothesis.* In L. T. Verhoeven, C. Elbro, & P. Reitsma (Eds.), Precursors of functional literacy, (pp. 67-86).

Perfetti, C., & Stafura, J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and Recovery From Garden Paths: An Eye-Tracking Study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 940-961. http://dx.doi.org/10.1037/0278-7393.24.4.940

Prat, C. S., & Just, M. A. (2011). Exploring the Neural Dynamics Underpinning Individual Differences in Sentence Comprehension. *Cerebral Cortex*, *21*(8), 1747–1760. https://doi.org/10.1093/cercor/bhq241

Rayner, K. (1978). Eye Movements in Reading and Information Processing. *Psychological Bulletin, 85*(3), 618-660. http://dx.doi.org/10.1037/0033-2909.85.3.618

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. https://doi.org/10.3758/BF03197692

Rayner, K., & Frazier, L. (1989). Selection Mechanisms in Reading Lexically Ambiguous Words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12.

Rayner, K., Pacht, J. M., & Duffy, S. A. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of memory and language*, *33*(4), 527-544.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The Effect of Plausibility on Eye Movements in Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290–1301. https://doi.org/10.1037/0278-7393.30.6.1290

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making Sense of Semantic Ambiguity: Semantic Competition in Lexical Access. *Journal of Memory and Language*, *46*(2), 245–266. https://doi.org/10.1006/jmla.2001.2810

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language*, *87*, 16–37. https://doi.org/10.1016/j.jml.2015.10.006

Rodd, J.M., Davis, M.H., & Johnsrude, I.S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex, 15*, 1261-1269.

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2010). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain and Language*, *115*(3), 182–188. https://doi.org/10.1016/j.bandl.2010.07.005

Rodd, J.M., Johnsrude, I.S., & Davis, M.H. (2012). Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. *Cerebral Cortex, 22*, 1761-73.

Rodd, J. M., Lopez Cutrin, B., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language*, *68*(2), 180–198. https://doi.org/10.1016/j.jml.2012.08.002

RStudio Team (2015). *RStudio: Integrated Development for R.* RStudio, Inc., Boston, MA. URL http://www.rstudio.com/

Schvaneveldt, R. W., Meyer, D. E., & Becker, C. A. (1976). Lexical ambiguity, semantic context and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2*(2), 243-256. http://dx.doi.org/10.1037/0096-1523.2.2.243

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, *14*(4), 489–537. https://doi.org/10.1016/0010-0285(82)90017-2

Sereno, S. C., Pacht, J. M., & Rayner, K. (1992). The Effect of Meaning Frequency on Processing Lexically Ambiguous Words: Evidence from Eye Fixations. *Psychological Science, 3*(5), 296–301. https://doi.org/10.1111/j.1467-9280.1992.tb00676.x

Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior*, *20*(1), 120–136. https://doi.org/10.1016/S0022-5371(81)90356-X

Simpson, G. B., & Burgess, C. (1985). Activation and Selection Processes in the Recognition of Ambiguous Words. *Journal of experimental psychology: Human perception and performance*, *11*(1), 28-39.

Simpson, G. B., & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, *30*(6), 627–643. https://doi.org/10.1016/0749-596X(91)90029-J

Stanovich, K E, & Cunningham, a E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, *20*(1), 51–68. https://doi.org/10.3758/BF03208254

Stanovich, Keith E., & West, R. F. (1989). Exposure to Print and Orthographic Processing. *Reading Research Quarterly*, *24*(4), 402. https://doi.org/10.2307/747605

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201–216. https://doi.org/10.3758/MC.36.1.201

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645–659. https://doi.org/10.1016/S0022-5371(79)90355-4

Tabor, W., & Hutchins, S. (2004). Evidence for Self-Organized Sentence Processing: Digging-In Effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 431–450. https://doi.org/10.1037/0278-7393.30.2.431

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of memory and language, 27*(3), 324-340. https://doi.org/10.1016/0749-596X(88)90058-7

Taylor, J. N., & Perfetti, C. A. (2016). Eye movements reveal readers' lexical quality and reading experience. *Reading and Writing, 29*(6), 1069-1103. https://doi.org/10.1007/s11145-015-9616-6

Traxler, M. J., & Tooley, K. M. (2007). Lexical mediation and context effects in sentence processing. *Brain Research*, *1146*, 59–74. https://doi.org/10.1016/j.brainres.2006.10.010

Trueswell, J. C., & Tanenhaus, M. K. (1994). Towards a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155-179). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Tunmer, W. E., & Chapman, J. W. (2012). The Simple View of Reading Redux: Vocabulary Knowledge and the Independent Components Hypothesis. *Journal of Learning Disabilities, 45*(5), 453-466. https://doi.org/10.1177/0022219411432685

Twilley, L. C., & Dixon, P. (2000). Meaning resolution processes for words: A parallel independent model. *Psychonomic Bulletin & Review*, *7*(1), 49–82. https://doi.org/10.3758/BF03210725

Twilley, L.C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition, 22*, 111-126.

Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(6), 683–692. https://doi.org/10.1016/S0022-5371(77)80027-3

Ulrich, R., & Miller, J. (1994). Effects of truncation on Reaction Time analysis. *Journal of Experimental Psychology: General*, *123*(1), 34-80.

Vitello, S., Warren, J. E., Devlin, J. T., & Rodd, J. M. (2014). Roles of frontal and temporal regions in reinterpreting semantically ambiguous sentences. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00530

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119–133. https://doi.org/10.1016/j.jml.2016.10.003

Vuong, L.C., & Martin, R.C. (2011). LIFG-based attentional control and the resolution of lexical ambiguities in sentence context. *Brain & Language, 116*, 22-32.

Wells, J., Christiansen, M., Race, D., Acheson, D., & Macdonald, M. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58*(2), 250–271. https://doi.org/10.1016/j.cogpsych.2008.08.002

Wollen, K.A., Cox, S.D., Coahran, M.M., Shea, D.S., & Kirby, R.F. (1980). Frequency of occurrence and concreteness ratings of homograph meanings. *Behavior Research Methods & Instrumentation, 12*, 8-15.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review, 23*(4), 1015-1027. https://doi.org/10.3758/s13423-015-0948-7

Appendix I

List of stimuli

Item	**Main Nouns**	**Sentence frame**	**Ambiguous word dominance score**[1]	
	(Coherent Unambiguous/ **Coherent Ambiguous**/ Anomalous	*(forward slashes separate regions-of-interest)*	**Single-word**	**Sentence-embedded**

	Unambiguous/ Anomalous Ambiguous)			
1	spurt/ **ace**/ prawn/ mule	The man knew that one more/MAIN NOUN/was enough to win the game of/tennis/against his/rival.	0.88	0.78
2	manuscript/ **appendix**/ cottage/ bowler	She asked about the/MAIN NOUN/and was told that it would be/translated/as soon/as possible.	0.58	0.79
3	guns/ **arms**/ bricks/ locks	The police kept hold of the criminals'/MAIN NOUN/after they had been/fired/one last/time.	0.83	0.75
4	town/ **ball**/ bird/ rule	Sally worried that the/MAIN NOUN/was going to be too/crowded/for her/liking.	0.84	0.84
5	boat/ **bank**/ knee/ press	The old man headed for the/MAIN NOUN/but he had a long way to/swim/to reach/it.	0.84	0.95
6	knife/ **bar**/ crowd/ court	The man found a/MAIN NOUN/but it was small and too/rusty/to be/used.	0.83	0.74
7	judge/ **board**/ desk/ rush	John hoped that the new/MAIN NOUN/would not take very long to/decide/this time/around.	0.44	0.53
8	kitten/ **boxer**/ fabric/ staple	The man knew that the/MAIN NOUN/would be in need of a good/veterinarian/as soon/as possible.	0.92	0.90
9	bus/ **branch**/ milk/ fit	The young woman couldn't use that/MAIN NOUN/because it was much/busier/than she/had expected.	0.92	0.95
10	hunt/ **bridge**/ fruit/ plane	Mary discussed the/MAIN NOUN/and said that it ought to be more/competitive/this coming/week.	0.88	0.95
11	berries/ **bulbs**/ sobs/ sharks	The caretaker preferred to store the/MAIN NOUN/inside a box in the old/greenhouse/until they/were needed.	0.52	0.79
12	lever/ **button**/ lawyer/ speaker	Karen knew that there was one last/MAIN	0.96	0.67

		NOUN/that she needed to/press/in order/to start.		
13	cliff/ **cape**/ ax/ horn	The woman saw the/MAIN NOUN/and thought that it looked very/rocky/from where/she stood.	0.96	0.95
14	fuse/ **chips**/ coal/ ray	The manager hoped that the/MAIN NOUN/he had ordered would be/compatible/this time/around.	0.54	0.65
15	roof/ **coat**/ pill/ shock	The man hoped that the new/MAIN NOUN/would not take long to/paint/early in/the morning.	0.92	0.80
16	nurse/ **cold**/ wall/ yard	Ted complained about the/MAIN NOUN/and said that he needed some/medicine/as soon/as possible.	0.80	0.67
17	lease/ **deed**/ harp/ mint	John thought about the/MAIN NOUN/and wondered whether it had been/signed/within the/deadline.	0.58	0.55
18	jail/ **dock**/ golf/ spit	The man was already at the/MAIN NOUN/but had to wait for the/magistrate/for another/hour.	0.83	0.89
19	parade/ **drill**/ cage/ pulse	The students liked the/MAIN NOUN/but thought it involved a lot more/marching/than they/had expected.	0.84	0.75
20	attention/ **interest**/ machine/ issue	The man expected that any/MAIN NOUN/he received would be/flattering/for him/tonight.	0.76	0.70
21	sheep/ **kids**/ tools/ scales	The mother knew that the/MAIN NOUN/were coming by the sound of their/hooves/getting/louder.	0.67	1.00
22	motor/ **lobby**/ breakfast/ relief	The plans meant that the new/MAIN NOUN/would be a lot more/powerful/in the/future.	0.75	0.79
23	map/ **log**/ zoo/ jet	They were surprised to find that the old/MAIN NOUN/was so easy to read/without much/help.	0.75	0.60

24	bomb/ **mine**/ sand/ page	The report said that the/MAIN NOUN/found by the children was not/armed/in the/end.	0.17	0.60
25	hinge/ **mouse**/ nest/ glare	The woman disliked the/MAIN NOUN/because of the way it/clicked/a bit/too loudly.	0.80	0.89
26	price/ **note**/ oil/ strike	The girl was surprised by the/MAIN NOUN/because it was much/higher/than she/had expected.	0.64	0.85
27	sandwich/ **nugget**/ cartoon/ landing	The man asked about the/MAIN NOUN/and was told that it was/chicken/this time/around.	0.39	0.53
28	dust/ **nut**/ wheat/ port	The boy wasn't paying attention to the/MAIN NOUN/when it fell off the/bolt/onto the/floor.	0.80	0.80
29	piano/ **organ**/ sergeant/ marble	The expert knew that the damaged/MAIN NOUN/would be quite difficult to/tune/after all/this time.	0.75	0.89
30	hedge/ **palm**/ cough/ pitch	The boy noticed the old man's/MAIN NOUN/and thought that it seemed more/wilted/than it/had before.	0.83	0.79
31	barn/ **pen**/ beer/ clutch	Frank compared the/MAIN NOUN/to the one he had bought for his/cattle/the previous/year.	1.00	1.00
32	carpet/ **pipe**/ ink/ spray	The boy saw that the/MAIN NOUN/was the same sort his father had/laid/the last/time.	0.65	0.53
33	church/ **plant**/ cook/ count	The newspaper reported that the/MAIN NOUN/had been very difficult to/build/from the/start.	0.96	0.90
34	necklace/ **poker**/ ravine/ pupil	Claire knew that this type of/MAIN NOUN/would probably be more/collectible/than all/the others.	0.88	0.85
35	computer/ **program**/	When Tom looked at the/MAIN NOUN/he saw that it now had	0.48	0.53

	kitchen/ capital	more/viruses/than ever/before.		
36	messages/ **records**/ battles/ matches	The students knew that the/MAIN NOUN/were so old they would be hard to/decipher/without a/lot of help.	0.74	0.70
37	dentist/ **ruler**/ ocean/ shower	The man hoped that the new/MAIN NOUN/would be a lot more/compassionate/than the/last.	0.88	0.79
38	memoir/ **scoop**/ spoon/ lash	The woman thought that such a big/MAIN NOUN/might be quite a challenge to/write/in such/a short time.	0.87	0.63
39	judgment/ **sentence**/ supper/ balance	When the old man heard the/MAIN NOUN/he thought it was too/lenient/under the/circumstances.	0.88	0.89
40	troops/ **shells**/ lakes/ knots	The men were told that all the/MAIN NOUN/had to be removed from the/arsenal/as quickly/as possible.	0.92	0.95
41	phone/ **staff**/ leader/ race	The man worried that a bigger/MAIN NOUN/would be more difficult to/hold/for too/long.	0.80	0.65
42	actor/ **star**/ fish/ ear	The news story on the/MAIN NOUN/included a picture and lots more/gossip/than ever/before.	0.92	0.74
43	disease/ **strain**/ cheese/ cap	The report said that the/MAIN NOUN/he was suffering from was not/contagious/for much/longer.	0.48	0.37
44	laws/ **tables**/ coffees/ posts	The professor disliked the/MAIN NOUN/because he thought they were too/confusing/to be/of use.	0.88	1.00
45	rain/ **tie**/ egg/ stock	Jim was worried about the/MAIN NOUN/and hoped it wouldn't ruin the/tournament/for everyone/involved.	0.92	1.00

46	previews/ **trailers**/ vases/ dressers	The boys saw one of the/MAIN NOUN/and asked to see the/movie/this coming/weekend.	0.60	0.50
47	poem/ **volume**/ aircraft/ cricket	The man explained why the/MAIN NOUN/was not going to be/published/the following/year.	0.88	0.84
48	police/ **watch**/ joke/ play	The man was annoyed that the/MAIN NOUN/had not been quite as/vigilant/as he/had hoped.	0.72	0.90

Appendix II

Results from analyses of accuracy and total sentence reading time

Dependent variable	**Predictor**						
Accuracy		*β*	*SE*	*z*	χ^2	*p*	
	Intercept	2.68	0.27	10.11			
	Ambiguity	3.5	0.26	13.26	110.62	<.001	*
	Vocabulary	0.03	0.21	0.15	0.02	.876	
	Print Exposure	0.03	0.2	0.16	0.02	.887	
	Ambiguity x Vocabulary	0.16	0.29	0.54	0.3	.587	
	Ambiguity x Print Exposure	0.03	0.29	0.12	0.01	.927	
Total sentence reading time		*β*	*SE*	*t*	χ^2	*p*	
	Intercept	3.54	0.01	263.37			
	Ambiguity	-0.05	0.004	-11.6	85.04	<.001	*
	Vocabulary	-0.02	0.01	-1.67	2.82	.093	
	Print Exposure	-0.03	0.01	-2.3	5.27	.022	*
	Ambiguity x Vocabulary	-0.01	0.01	-1.17	1.4	.236	
	Ambiguity x Print Exposure	0.003	0.01	0.56	0.32	.56	

**Statistically significant at* α = *.05*

App. II Table 1. **Results from linear and logit mixed effect model comparisons for accuracy and total sentence reading time.** Models were fitted to data from the Coherent Ambiguous and the Coherent Unambiguous condition only; for total sentence reading time, only correct trials were included. Fixed effect parameter estimates, standard error (SE), z- and t-values, and Chi-squared and *p*-values from likelihood ratio tests are reported. Maximal models were structured as follows: Dependent variable ~ 1 + Ambiguity + Vocabulary + Print Exposure + Ambiguity : Vocabulary + Ambiguity : Print Exposure + (1 + Vocabulary + Print Exposure | Items) + (1 + Ambiguity | Subjects)

Results from analyses of eye-tracking measures in the Main Noun region

Dependent variable	Predictor	*β*	*SE*	*t*	χ^2	*p*	
First-fixation duration							
	Intercept	2.35	0.01	420.44			
	Ambiguity	-0.01	0.01	-2.04	4.16	.042	[1]
	Vocabulary	-0.004	0.01	-0.56	0.33	.568	
	Print Exposure	-0.01	0.01	-1.19	1.45	.229	
	Ambiguity x Vocabulary	-0.01	0.01	-1.11	1.22	.269	
	Ambiguity x Print Exposure	0.003	0.01	0.5	0.24	.623	
Gaze duration							
	Intercept	2.38	0.01	357.92			
	Ambiguity	-0.01	0.01	-1.69	2.85	.091	
	Vocabulary	-0.01	0.01	-0.76	0.59	.443	
	Print Exposure	-0.02	0.01	-2.44	5.91	.015	[1]
	Ambiguity x Vocabulary	-0.001	0.01	-0.16	0.03	.873	
	Ambiguity x Print Exposure	-0.003	0.01	-0.4	0.16	.691	
Go-past time							
	Intercept	2.45	0.01	269.98			
	Ambiguity	-0.01	0.01	-1.08	1.17	.28	
	Vocabulary	-0.01	0.01	-1.18	1.41	.236	
	Print Exposure	-0.03	0.01	-3.0	8.84	.003	*
	Ambiguity x Vocabulary	-0.01	0.01	-1.3	1.68	.195	
	Ambiguity x Print Exposure	0.002	0.01	0.22	0.05	.831	
Second-pass reading time							
	Intercept	2.38	0.01	250.59			
	Ambiguity	-0.05	0.02	-2.72	7.22	.007	*
	Vocabulary	-0.0004	0.01	-0.04	0.002	.961	
	Print Exposure	0.002	0.01	0.13	0.02	.89	
	Ambiguity x Vocabulary	-0.01	0.02	-0.63	0.42	.519	
	Ambiguity x Print Exposure	-0.02	0.02	-0.69	0.5	.478	

		β	*SE*	*z*	χ^2	*p*	
Regressions out							
	Intercept	-1.91	0.1	-19.9			
	Ambiguity	-0.17	0.12	1.46	2.09	.148	
	Vocabulary	-0.05	0.1	-0.55	0.3	.585	
	Print Exposure	-0.17	0.1	-1.65	2.69	.101	
	Ambiguity x Vocabulary	-0.24	0.12	-2.0	3.9	.048	[1]
	Ambiguity x Print Exposure	0.27	0.14	2.0	3.96	.047	[1]

**Statistically significant at Bonferroni-corrected* $\alpha = .01$

[1] *Statistically significant at uncorrected* $\alpha = .05$

App. II Table 2. **Results from linear and logit mixed effect model comparisons for eye tracking measures in the Main Noun region.** Models for first-pass measures (first-fixation duration, gaze duration, go-past time and regressions out) were fitted to data from the combined ambiguous

(Coherent Ambiguous and Anomalous Ambiguous) and the combined unambiguous conditions (Coherent Unambiguous and Anomalous Unambiguous). Models for second-pass reading time were fitted to data from the Coherent Ambiguous and the Coherent Unambiguous condition only. Separate models were fitted for each dependent variable. Fixed effect parameter estimates, standard error (SE), z- and t-values, and Chi-squared and *p*-values from likelihood ratio tests are reported. Maximal models were structured as follows: Dependent variable ~ 1 + Ambiguity + Vocabulary + Print Exposure + Ambiguity : Vocabulary + Ambiguity : Print Exposure + (1 + Vocabulary + Print Exposure | Items) + (1 + Ambiguity | Subjects). For first-pass and second-pass reading times, correlations between random effects were removed from the model due to convergence issues.

Model	**Dependent variable**	**Predictor**	*β*	*SE*	*z*	χ^2	*p*
Ambiguous trials	*Regressions out*						
		Intercept	-2.16	0.23	-9.32		
		Vocabulary	-0.04	0.18	-0.2	0.04	.844
		Print Exposure	-0.47	0.24	-2.0	4.6	.032[1]
Unambiguous trials	*Regressions out*						
		Intercept	-2.25	0.21	-10.7		
		Vocabulary	-0.29	0.19	-1.56	2.5	.114
		Print Exposure	0.05	0.2	0.27	0.07	.787

[*] *Statistically significant at Bonferroni-corrected* α = *.025*

[1] *Statistically significant at uncorrected* α = *.05*

App. II Table 3. **Pairwise comparisons of regressions out of the Main Noun region.** Models were fitted to data from the combined ambiguous (Coherent Ambiguous and Anomalous Ambiguous) and the combined unambiguous conditions (Coherent Unambiguous and Anomalous Unambiguous). Separate models were fitted for the Amb and Unamb conditions. Fixed effect parameter estimates, standard error (SE), z- and t-values, and Chi-squared and *p*-values from likelihood ratio tests are reported. Maximal models were structured as follows: Regressions out ~ 1 + Vocabulary + Print Exposure + (1 + Vocabulary + Print Exposure | Items) + (1 | Subjects)

Results from analyses of eye-tracking measures in the Coherence cue region

Dependent variable	**Predictor**	*β*	*SE*	*t*	χ^2	*p*
First-fixation duration						
	Intercept	2.36	0.01	327.33		
	Ambiguity	-0.002	0.01	-0.33	0.12	.734
	Vocabulary	-0.004	0.01	-0.64	0.43	.515
	Print Exposure	-0.003	0.01	-0.54	0.3	.582
	Ambiguity x Vocabulary	0.02	0.01	1.9	3.67	.055
	Ambiguity x Print Exposure	-0.01	0.01	-0.74	0.56	.454
Gaze duration						

	Intercept	2.4	0.01	234.413		
	Ambiguity	-0.002	0.01	-0.22	0.05	.828
	Vocabulary	-0.01	0.01	-1.37	1.9	.168
	Print Exposure	-0.01	0.01	-1.7	2.89	.089
	Ambiguity x Vocabulary	-0.01	0.01	-0.77	0.63	.428
	Ambiguity x Print Exposure	0.01	0.01	0.6	0.36	.55
Go-past time						
	Intercept	2.46	0.01	202.82		
	Ambiguity	-0.02	0.01	-1.64	2.74	.098
	Vocabulary	-0.02	0.01	-2.06	4.24	.04[1]
	Print Exposure	-0.02	0.01	-1.78	3.16	.076
	Ambiguity x Vocabulary	-0.01	0.01	-1.0	1.03	.31
	Ambiguity x Print Exposure	0.01	0.01	0.58	0.36	.552
Second-pass reading time						
	Intercept	2.35	0.02	152.94		
	Ambiguity	-0.05	0.02	-2.2	4.79	.029[1]
	Vocabulary	-0.01	0.01	-0.91	0.81	.367
	Print Exposure	-0.002	0.02	-0.15	0.02	.885
	Ambiguity x Vocabulary	-0.002	0.03	-0.09	0.005	.944
	Ambiguity x Print Exposure	0.01	0.03	0.29	0.07	.792

		β	*SE*	*z*	χ^2	*p*
Regressions out						
	Intercept	-2.21	0.15	-14.47		
	Ambiguity	-0.21	0.21	-1.01	1.21	.271
	Vocabulary	-0.01	0.14	-0.06	0.004	.95
	Print Exposure	-0.29	0.17	-1.75	3.19	.074
	Ambiguity x Vocabulary	-0.15	0.2	-0.77	0.59	.442
	Ambiguity x Print Exposure	0.18	0.22	0.83	0.68	.409

[1] *Statistically significant at uncorrected* $\alpha = .05$

App. II Table 4. **Results from linear and logit mixed effect model comparisons for eye tracking measures in the Coherence cue region.** Models were fitted to data from the Coherent Ambiguous and the Coherent Unambiguous condition only. One item was removed from analyses due to issues with sentence region assignment. Separate models were fitted for each dependent variable. Fixed effect parameter estimates, standard error (SE), z- and t-values, and Chi-squared and *p*-values from likelihood ratio tests are reported. Maximal models were structured as follows: Dependent variable ~ 1 + Ambiguity + Vocabulary + Print Exposure + Ambiguity : Vocabulary + Ambiguity : Print Exposure + (1 + Vocabulary + Print Exposure | Items) + (1 + Ambiguity | Subjects). For first-pass reading times, correlations between random effects were removed from the model due to convergence issues.

Results from analyses of eye-tracking measures in the Spill-over region

Dependent variable	**Predictor**	*β*	*SE*	*t*	χ^2	*p*	
First-fixation duration							
	Intercept	2.4	0.01	292.45			
	Ambiguity	0.001	0.01	0.14	0.02	.879	
	Vocabulary	0.0001	0.01	0.01	0	.996	
	Print Exposure	-0.004	0.01	-0.52	0.28	.598	
	Ambiguity x Vocabulary	-0.02	0.01	-1.65	2.77	.096	
	Ambiguity x Print Exposure	-0.01	0.01	1.22	1.51	.219	
Gaze duration							
	Intercept	2.5	0.02	168.71			
	Ambiguity	-0.01	0.01	-1.26	1.59	.207	
	Vocabulary	-0.003	0.01	-0.25	0.06	.803	
	Print Exposure	-0.02	0.01	-1.33	1.79	.181	
	Ambiguity x Vocabulary	-0.02	0.01	-1.93	3.7	.054	
	Ambiguity x Print Exposure	0.02	0.01	2.02	4.08	.043	1
Go-past time							
	Intercept	2.7	0.03	92.52			
	Ambiguity	-0.07	0.01	-4.86	23.04	<.001	*
	Vocabulary	0.003	0.02	0.16	0.03	.868	
	Print Exposure	-0.03	0.02	-1.5	2.27	.132	
	Ambiguity x Vocabulary	-0.03	0.02	-1.91	3.63	.057	
	Ambiguity x Print Exposure	0.03	0.02	1.66	2.75	.098	
Second-pass reading time							
	Ambiguity	0.01	0.02	0.59	0.36	.55	
	Vocabulary	0.01	0.02	0.61	0.4	.528	
	Print Exposure	-0.04	0.02	-2.54	6.37	.012	1
	Ambiguity x Vocabulary	-0.01	0.02	-0.59	0.35	.557	
	Ambiguity x Print Exposure	0.03	0.02	1.23	1.49	.223	
		β	*SE*	*z*	χ^2	*p*	
Regressions out							
	Intercept	-0.81	0.19	-4.32			
	Ambiguity	-0.42	0.13	-3.32	10.63	.001	*
	Vocabulary	0.07	0.13	0.49	0.24	.624	
	Print Exposure	-0.09	0.13	-0.7	0.48	.488	

Ambiguity x Vocabulary	-0.3	0.15	-2.06	4.15	.042[1]
Ambiguity x Print Exposure	0.19	0.15	1.29	1.63	.202

**Statistically significant at Bonferroni-corrected* α = *.01*

[1] *Statistically significant at uncorrected* α = *.05*

App. II Table 5. **Results from linear and logit mixed effect model comparisons for eye tracking measures in the Spill-over region.** Models were fitted to data from the Coherent Ambiguous and the Coherent Unambiguous condition only. One item was removed from analyses due to issues with sentence region assignment. Separate models were fitted for each dependent variable. Fixed effect parameter estimates, standard error (SE), z- and t-values, and Chi-squared and *p*-values from likelihood ratio tests are reported. Maximal models were structured as follows: Dependent variable ~ 1 + Ambiguity + Vocabulary + Print Exposure + Ambiguity : Vocabulary + Ambiguity : Print Exposure + (1 + Vocabulary + Print Exposure | Items) + (1 + Ambiguity | Subjects)

Model	**Dependent variable**	**Predictor**	*β*	*SE*	*t*	χ^2	*p*
Coherent Ambiguous trials	*Gaze duration*						
		Intercept	2.51	0.02	165.31		
		Vocabulary	0.01	0.02	0.42	0.18	.669
		Print Exposure	-0.03	0.02	-1.89	3.58	.058
Coherent Unambiguous trials	*Gaze duration*						
		Intercept	2.50	0.02	152.69		
		Vocabulary	-0.02	0.01	-1.11	1.26	.261
		Print Exposure	-0.01	0.01	-0.35	0.13	.724

**Statistically significant at Bonferroni-corrected* α = *.025*

[1] *Statistically significant at uncorrected* α = *.05*

App. II Table 6. **Pairwise comparisons of gaze duration in the Spill-over region.** Models were fitted to data from the Coherent Ambiguous and the Coherent Unambiguous condition only. One item was removed from analyses due to issues with sentence region assignment. Separate models were fitted for the Coherent Ambiguous and Coherent Unambiguous conditions. Fixed effect parameter estimates, standard error (SE), z- and t-values, and Chi-squared and *p*-values from likelihood ratio tests are reported. Maximal models were structured as follows: Gaze duration ~ 1 + Vocabulary + Print Exposure + (1 + Vocabulary + Print Exposure | Items) + (1 | Subjects)

Model	**Dependent variable**	**Predictor**	*β*	*SE*	*z*	χ^2	*p*
Coherent Ambiguous trials	*Regression out*						
		Intercept	-0.57	0.2	-2.83		
		Vocabulary	0.23	0.15	1.46	2.12	.145
		Print Exposure	-0.17	0.15	-1.09	1.17	.279
Coherent Unambiguous trials	*Regression out*						

Intercept	-0.99	0.18	-5.53		
Vocabulary	-0.07	0.14	-0.52	0.26	.608
Print Exposure	-0.003	0.14	-0.02	0	.983

Statistically significant at Bonferroni-corrected α = .01

[1] *Statistically significant at uncorrected* α = *.05*

App. II Table 7. **Pairwise comparisons of regressions out of the Spill-over region.** Models were fitted to data from the Coherent Ambiguous and the Coherent Unambiguous condition only. One item was removed from analyses due to issues with sentence region assignment. Separate models were fitted for the Coherent Ambiguous and Coherent Unambiguous conditions. Fixed effect parameter estimates, standard error (SE), z- and t-values, and Chi-squared and *p*-values from likelihood ratio tests are reported. Maximal models were structured as follows: Regressions out ~ 1 + Vocabulary + Print Exposure + (1 + Vocabulary + Print Exposure | Items) + (1 | Subjects)

Results from analyses of eye-tracking measures in the Wrap-up region

Dependent variable	**Predictor**	*β*	*SE*	*t*	χ^2	*p*	
First-fixation duration							
	Intercept	2.37	0.01	248.11			
	Ambiguity	-0.01	0.01	-0.48	0.23	.634	
	Vocabulary	0.01	0.01	0.71	0.51	.476	
	Print Exposure	-0.02	0.01	-1.85	3.34	.068	
	Ambiguity x Vocabulary	-0.01	0.01	-0.47	0.22	.637	
	Ambiguity x Print Exposure	0.0004	0.01	0.04	0.001	.976	
Gaze duration							
	Intercept	2.42	0.02	145.79			
	Ambiguity	-0.02	0.01	-1.67	2.74	.098	
	Vocabulary	0.01	0.02	0.75	0.56	.453	
	Print Exposure	-0.04	0.02	-2.81	7.28	.007	*
	Ambiguity x Vocabulary	-0.002	0.02	-0.14	0.02	.894	
	Ambiguity x Print Exposure	0.01	0.02	0.56	0.3	.582	
Go-past time							
	Intercept	2.86	0.03	112.54			
	Ambiguity	-0.18	0.02	-9.89	62.88	<.001	*
	Vocabulary	-0.004	0.02	-0.15	0.02	.88	
	Print Exposure	-0.03	0.02	-1.4	2	.158	
	Ambiguity x Vocabulary	-0.01	0.02	-0.29	0.09	.763	
	Ambiguity x Print Exposure	-0.004	0.02	-0.21	0.04	.837	
Second-pass reading time							
	Intercept	2.42	0.02	116.92			
	Ambiguity	-0.01	0.04	-0.22	0.03	.871	
	Vocabulary	0.03	0.02	1.67	2.79	.095	
	Print Exposure	-0.05	0.03	-1.84	2.66	.103	
	Ambiguity x Vocabulary	0.01	0.04	0.15	0.03	.863	
	Ambiguity x Print Exposure	-0.02	0.05	-0.43	0.18	.674	

		β	*SE*	z	χ^2	p	
Regressions out							
	Intercept	1.21	0.18	6.82			
	Ambiguity	-0.77	0.18	-4.21	17.81	<.001	*
	Vocabulary	0.03	0.17	0.14	0.02	0.888	
	Print Exposure	0.06	0.17	0.32	0.1	0.75	
	Ambiguity x Vocabulary	-0.16	0.19	-0.86	0.72	0.397	
	Ambiguity x Print Exposure	-0.01	0.19	-0.06	0.004	0.95	

[*] *Statistically significant at Bonferroni-corrected* α = *.01*

[1] *Statistically significant at uncorrected* α = *.05*

App. II Table 8. **Results from linear and logit mixed effect model comparisons for eye tracking measures in the Wrap-up region.** Models were fitted to data from the Coherent Ambiguous and the Coherent Unambiguous condition only. Separate models were fitted for each dependent variable. Fixed effect parameter estimates, standard error (SE), z- and t-values, and Chi-squared and *p*-values from likelihood ratio tests are reported. Maximal models were structured as follows: Dependent variable ~ 1 + Ambiguity + Vocabulary + Print Exposure + Ambiguity : Vocabulary + Ambiguity : Print Exposure + (1 + Vocabulary + Print Exposure | Items) + (1 + Ambiguity | Subjects)

Model	**Predictor**	**VIF**
Accuracy		
	Vocabulary	1.99
	Print exposure	2.03
	Ambiguity x Vocabulary	2.15
	Ambiguity x Print exposure	2.18
Total sentence reading time		
	Vocabulary	1.46
	Print exposure	1.46
	Ambiguity x Vocabulary	1.44
	Ambiguity x Print exposure	1.44
Main Noun region		
First-fixation duration		
	Vocabulary	1.4
	Print exposure	1.4
	Ambiguity x Vocabulary	1.41
	Ambiguity x Print exposure	1.41
Gaze duration		
	Vocabulary	1.42
	Print exposure	1.41
	Ambiguity x Vocabulary	1.42
	Ambiguity x Print exposure	1.42
Go-past time		
	Vocabulary	1.36
	Print exposure	1.36
	Ambiguity x Vocabulary	1.4

Ambiguity x Print exposure	1.4
Second-pass reading time	
Vocabulary	1.32
Print exposure	1.33
Ambiguity x Vocabulary	1.41
Ambiguity x Print exposure	1.42
Regressions out	
Vocabulary	1.38
Print exposure	1.39
Ambiguity x Vocabulary	1.46
Ambiguity x Print exposure	1.5
Coherence cue region	
First-fixation duration	
Vocabulary	1.4
Print exposure	1.41
Ambiguity x Vocabulary	1.38
Ambiguity x Print exposure	1.39
Gaze duration	
Vocabulary	1.35
Print exposure	1.35
Ambiguity x Vocabulary	1.42
Ambiguity x Print exposure	1.42
Go-past time	
Vocabulary	1.38
Print exposure	1.38
Ambiguity x Vocabulary	1.44
Ambiguity x Print exposure	1.43
Second-pass reading time	
Vocabulary	1.46
Print exposure	1.44
Ambiguity x Vocabulary	1.47
Ambiguity x Print exposure	1.45
Regressions out	
Vocabulary	1.38
Print exposure	1.38
Ambiguity x Vocabulary	1.41
Ambiguity x Print exposure	1.45
Spill-over region	
First-fixation duration	
Vocabulary	1.38
Print exposure	1.38
Ambiguity x Vocabulary	1.37
Ambiguity x Print exposure	1.37
Gaze duration	
Vocabulary	1.4
Print exposure	1.4
Ambiguity x Vocabulary	1.39

		Ambiguity x Print exposure	1.39
	Go-past time		
		Vocabulary	1.49
		Print exposure	1.49
		Ambiguity x Vocabulary	1.48
		Ambiguity x Print exposure	1.48
	Second-pass reading time		
		Vocabulary	1.36
		Print exposure	1.35
		Ambiguity x Vocabulary	1.39
		Ambiguity x Print exposure	1.38
	Regressions out		
		Vocabulary	1.39
		Print exposure	1.39
		Ambiguity x Vocabulary	1.38
		Ambiguity x Print exposure	1.38
Wrap-up region			
	First-fixation duration		
		Vocabulary	1.46
		Print exposure	1.46
		Ambiguity x Vocabulary	1.4
		Ambiguity x Print exposure	1.4
	Gaze duration		
		Vocabulary	1.43
		Print exposure	1.43
		Ambiguity x Vocabulary	1.39
		Ambiguity x Print exposure	1.39
	Go-past time		
		Vocabulary	1.42
		Print exposure	1.41
		Ambiguity x Vocabulary	1.39
		Ambiguity x Print exposure	1.39
	Second-pass reading time		
		Vocabulary	1.13
		Print exposure	1.13
		Ambiguity x Vocabulary	1.31
		Ambiguity x Print exposure	1.37
	Regressions out		
		Vocabulary	1.42
		Print exposure	1.41
		Ambiguity x Vocabulary	1.41
		Ambiguity x Print exposure	1.39
Predicting accuracy from regressions out of the Spill-over region			
		Vocabulary	2.47
		Print exposure	2.24
		Ambiguity x Vocabulary	3.01
		Ambiguity x Print exposure	2.64

Regressions x Vocabulary	2.39
Regressions x Print exposure	2.24
Ambiguity x Regressions x Vocabulary	2.8
Ambiguity x Regressions x Print exposure	2.64

App. II Table 9. **Variance inflation factors (VIFs) for the accuracy, total sentence reading time, and eye-tracking measure models.**