

Presentation format affects the behavioural and neural processing costs of sentence reinterpretation

Lena M. Blott, Jennifer M. Rodd & Jane E. Warren

To cite this article: Lena M. Blott, Jennifer M. Rodd & Jane E. Warren (2023): Presentation format affects the behavioural and neural processing costs of sentence reinterpretation, Language, Cognition and Neuroscience, DOI: [10.1080/23273798.2022.2159992](https://doi.org/10.1080/23273798.2022.2159992)

To link to this article: <https://doi.org/10.1080/23273798.2022.2159992>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 04 Jan 2023.



[Submit your article to this journal](#)



Article views: 138




[View related articles](#)



[View Crossmark data](#)

Presentation format affects the behavioural and neural processing costs of sentence reinterpretation

Lena M. Blott , Jennifer M. Rodd and Jane E. Warren

Division of Psychology and Language Sciences, University College London, London, United Kingdom

ABSTRACT

Although listening to speech and reading text rely on different lower-order cognitive and neural processes, much of the literature on higher-order comprehension assumes engagement of a common conceptual-semantic system which is unaffected by input modality. However, few studies have tested this assumption directly. Moreover, many neuroimaging studies of reading present sentences in an artificial, cognitively demanding word-by-word rapid serial visual presentation (RSVP) format. We report behavioural and fMRI experiments investigating whether presentation format (Spoken, Written, or RSVP) modulates commonly reported behavioural and neural costs associated with reinterpretation of sentences that contain lexical ambiguities. Reinterpretation-related processing costs were exaggerated in the RSVP format, both for response times on a behavioural task and neural activation in left inferior frontal gyrus. Presentation format can interact with higher-order language processes in complex ways, and we urge language researchers to carefully consider the role of presentation format in study design and interpretation of research findings.

ARTICLE HISTORY

Received 24 February 2022
Accepted 7 December 2022

KEYWORDS

Ambiguity resolution;
language comprehension;
modality; inferior frontal
gyrus


Introduction

A tacit assumption underlying much of the neuroimaging and behavioural literature on language comprehension is that higher-level language processes, such as the resolution of lexical-semantic ambiguity, operate in a similar manner irrespective of the format in which the linguistic material is presented. While listening to speech and reading written material engage different low-level, unimodal sensory regions, language processing pathways are assumed to converge onto a common amodal conceptual-semantic system that is engaged during sentence processing in any given input modality (Binder et al., 2009; Hickok & Poeppel, 2004; Mesulam, 1998). Functional neuroimaging studies comparing the spoken and written input modalities have supported this assumption by demonstrating convergent activation for auditory and visual language in multimodal cortical regions in the temporal and left inferior frontal lobes (Marinkovic et al., 2003; Spitsyna et al., 2006). Based on the assumption that extracting meaning from spoken and written material engages the same cognitive and neural processes, most neuroimaging studies investigating semantic aspects of sentence processing have presented stimuli in a single

modality only (Crinion et al., 2003; Fedorenko et al., 2011; Rodd et al., 2005; Rodd et al., 2012; Rodd, Johnsrude, et al., 2010; Rodd, Longe, et al., 2010; Vitello et al., 2014; Zhu et al., 2012). Similarly, much behavioural work on higher-level language processing has been conducted on either spoken or written sentences alone, under the assumption that conclusions from either presentation modality apply to language processing more generally.

The evidence base for many psycholinguistic phenomena has relied on aggregation of data from single-modality studies of spoken and written language processing. For example, the rich literature on lexical-semantic ambiguity resolution suggests that, in comparison to matched unambiguous sentences, ambiguous sentences tend to be more difficult to understand, take longer to process, and are associated with additional neural activity in typical language regions, most notably left inferior frontal gyrus (IFG) (Duffy et al., 1988; Kambe et al., 2001; Rodd, 2020; Rodd et al., 2012; Simpson, 1981; Swinney, 1979; Twilley & Dixon, 2000; for a recent review see Rodd, 2020). Crucially, the evidence for these behavioural and neural “processing costs” associated with lexical-semantic ambiguity

CONTACT Lena M Blott  lena.blott.12@ucl.ac.uk

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23273798.2022.2159992>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

resolution comes from studies that presented sentences in either spoken or in written format (Mason & Just, 2007; Rodd et al., 2005; Rodd et al., 2012; Vitello & Rodd, 2015; Zempleni et al., 2007). However, no study to date has attempted to test the assumption that these ambiguity-related processing costs are unaffected by input modality by directly comparing the magnitude of processing costs across different language presentation formats.

A fundamental difference between the spoken and written format is that written material remains visibly accessible to the reader. In contrast to listeners, readers can control the rate and order of language input themselves, including whether they return to earlier parts of a sentence during reading to facilitate comprehension. This serves to mitigate demands on memory processes, but also hampers use of the whole-sentence written format for studies which rely on time-locking cognitive processes or neural responses to specific words or phrases within a sentence. Neuroimaging researchers in particular, therefore, have often relied on alternative visual presentation techniques which standardise temporal exposure to written sentence content across participants, such as phrase-by-phrase or word-by-word presentation (e.g. Fedorenko et al., 2011; Hoenig & Scheef, 2009; Zhu et al., 2012).

One popular presentation technique is rapid serial visual presentation (RSVP), where the words that make up a sentence are presented individually in quick succession in the centre of the visual field, minimising readers' eye movements and enabling measurement of responses time-locked to specific sentence elements. RSVP has been used in neuroimaging studies of high-level language processing under the assumption that, although low-level processing may differ from natural whole-sentence reading, high-level comprehension processes remain the same. However, RSVP is a highly artificial sentence format that is unfamiliar to most research participants, and its use has been associated with poorer text comprehension in comparison to natural whole-sentence reading (Masson, 1983; Schotter et al., 2014; Waters & Caplan, 1996; Wlotko & Federmeier, 2015). As RSVP is widely used in neuroimaging to enable modelling of event-related haemodynamic responses, an understanding of potential interactions between this presentation format and complex comprehension processes (such as those involved in ambiguity resolution) is essential. This is particularly the case because evidence of leftIFG involvement in top-down control of language processing and working memory (e.g. Bookheimer, 2002; Curtis & D'Esposito, 2003; Kaan & Swaab, 2003; Nozari & Thompson-Schill, 2016; Postle, 2006; Thompson-Schill et al., 1997) raises the possibility that previous findings of leftIFG activation in

response to RSVP stimuli could have conflated higher-level language-related activation with activation related specifically to the processing demands of the presentation format.

It follows that a key question for the field is whether, when investigating higher-order language processes, it matters, not only *which* stimuli are presented, but *how* stimuli are presented to participants. To our knowledge, only a single previous neuroimaging study has directly compared the effects of presentation format on behavioural and haemodynamic measures during higher-level language processing. Lee and Newman (2009) used an explicit comprehension task to investigate the effects of syntactic complexity during reading of sentences presented in RSVP and whole-sentence format, and found that the presentation format influenced both behavioural and neural measures of sentence processing. The adverse effects of syntactic complexity on comprehension accuracy were enhanced when sentences were presented in RSVP format. Presentation format also modulated haemodynamic responses to syntactic complexity, but in the opposite direction. Greater syntactic complexity effects were observed in left IFG during whole-sentence reading compared to reading RSVP sentences, in the context of an explicit comprehension task; though the authors attributed this finding to the additional neural processing required for baseline non-complex sentences in RSVP format.

The above findings demonstrate that how linguistic information is presented does indeed influence behavioural and neural measures of one type of higher-order comprehension process, the successful resolution of syntactic complexity. An important outstanding question is whether such effects occur in higher-level sentence comprehension processes beyond syntactic processing. Additionally, it is unclear whether the previously demonstrated processing differences between RSVP and whole-sentence text relate to the transience and temporal constraints of the RSVP format, or to its artificiality as a mode of language presentation.

In the present study, we investigated the impact of language presentation format on behavioural and functional neuroimaging measures of a different higher-order comprehension process, the resolution of lexical-semantic ambiguity in sentence context. The two experiments presented in this study used sentences that contained ambiguous words with a more frequent ("dominant") meaning and a less frequent ("subordinate") meaning, e.g. "bark". In the absence of informative prior context, such ambiguities are initially resolved towards the dominant meaning (Rodd, 2018; Twilley & Dixon, 2000). In each sentence, a single ambiguous word was presented within a neutral sentential context

until the final word in the sentence, which constrained the context in favour of the subordinate meaning of the ambiguous word (e.g. “The woman thought that the bark must have come from her neighbour’s willow.”). Comprehenders were therefore initially led down a “garden path” (selecting the dominant meaning), and needed to revise their interpretation upon encounter of the final word. This stimulus design had the advantage of allowing partial temporal separation of lower-level processing during sentence presentation from the higher-order processes of revising and reinterpreting sentence meaning that were triggered by the sentence-final word.

Ambiguous sentences were presented in three formats (spoken, whole-sentence written, and RSVP). This design allowed the RSVP format, a mode of sentence presentation rarely encountered in daily life and therefore likely to be associated with greater attentional and working memory demands, to be contrasted with naturalistic and familiar spoken and whole-sentence written formats. The use of spoken sentences provided a transient, externally-paced but naturalistic language presentation format for comparison with RSVP and whole-sentence reading. We did not attempt to equalise the processing demands of each format, but instead utilised each as they are typically presented in psycholinguistic and neurolinguistic experiments. Our aim was to use separate behavioural (Experiment 1) and fMRI (Experiment 2) paradigms to investigate whether the idiosyncratic processing demands of these typical presentation formats modulate the well-established behavioural and neural costs of reinterpreting sentence meaning to resolve semantic ambiguity, either in the context of an explicit comprehension task or during implicit language comprehension.

Experiment 1: behavioural processing costs of ambiguity resolution

Experiment 1 compared the behavioural processing costs associated with ambiguity resolution in the Spoken, Written, and RSVP presentation formats. Participants performed an explicit Meaning Coherence Judgement task (“Sense” vs “Nonsense”) on ambiguous and matched unambiguous sentences, as well as semantically anomalous filler sentences, presented in Spoken, whole-sentence Written and RSVP formats. A correct decision for an Ambiguous sentence (“Sense”) required successful reinterpretation of lexical-semantic ambiguity. Accuracy rates and response times for correct sentence judgements were compared across the three presentation formats. Methods and hypotheses for

Experiment 1 were preregistered prior to data analysis (<https://osf.io/9ryxu/>).

In line with the previous literature on the behavioural processing costs associated with lexical-semantic ambiguity resolution (e.g. Christianson et al., 2001; Duffy et al., 1988; Kambe et al., 2001), we expected that ambiguous sentences, which required sentence reinterpretation, would be associated with lower accuracy and longer response time on correct trials compared to the unambiguous sentences. We predicted that these ambiguity effects on accuracy and response times in the Meaning Coherence Judgement task would be exaggerated in the RSVP format compared to the Spoken and Written conditions (see e.g. Lee & Newman, 2009).

Method

Participants

A total sample size of 108 participant was targeted (see pre-registration document) based on a sample size calculation with G*Power prior to data collection. This calculation was – conservatively – based on a 2×3 repeated measures, within-factors ANOVA design to ensure sufficient statistical power to detect a small effect (.1) at a significance threshold of $\alpha = .05$. The analyses included data from 108 speakers of British English (71 female, aged 18–35, $M_{Age} = 26.0 \pm 4.9$ years), who were residing in the UK at the time of the study, had normal or corrected-to-normal vision, and had no history of significant neurological or developmental language disorder. All participants either self-identified as native speakers of British English ($n = 104$), or had acquired the language before age 11 ($n = 4$). Sixty-seven of our participants had completed a university degree, while 41 had not completed university-level education. An additional 16 participants were excluded from analyses due to procedural or technical difficulties with data collection.

Participants were recruited via the web-based recruitment service Prolific (www.prolific.ac; Damer & Bradley, 2014), and completed the study remotely in their own time. Ethical approval for this study was obtained from the UCL Department of Language and Cognition Ethics Chair as part of a larger research programme. Written informed consent was obtained from all participants, and they were remunerated for participation at standard UCL rates.

Materials

Stimuli for Experiments 1 and 2 were based on a set of 90 sentences, each containing a different ambiguous key noun (e.g. “bark”, see Ambiguous example in Table 1). The sentential context before the ambiguous key noun

Table 1. Example stimuli in the two coherent conditions (Ambiguous, Unambiguous) and the anomalous filler conditions.

Condition	Example sentence
Ambiguous	The woman thought that the bark must have come from her neighbour's willow.
Unambiguous	The woman thought that the fumes must have come from her neighbour's boiler.
Anomalous Ambiguous	The woman thought that such a big scoop might be quite a challenge to sharpen.
Anomalous Unambiguous	The woman thought that such a big quest might be quite a challenge to wash.

was neutral, and therefore compatible with dominant and subordinate meanings of the ambiguous word. The final word disambiguated the key noun to its less frequent, subordinate meaning (e.g. "willow"). Key noun and final word were separated by 4–8 intervening words, to allow initial key noun meaning selection to take place prior to encountering the final word. Since key nouns were biased in their relative meaning frequency, it was assumed that comprehenders would initially select the dominant meaning, and would then need to reinterpret the sentence after processing the disambiguating final word (e.g. Twilley & Dixon, 2000).

Table 2. Stimulus characteristics. Sentences in the Ambiguous condition contained an ambiguous key noun (e.g. "bark") that was disambiguated towards its subordinate meaning by the final word (e.g. "willow"). Control sentences in the Unambiguous condition contained a frequency-matched and length-matched unambiguous key noun and final word. Means (SD) are reported for psycholinguistic characteristics of the stimuli.

Characteristic	Condition	
	Ambiguous	Unambiguous
N	90	90
Number of syllables	16.42 (1.5)	16.42 (1.5)
Number of words	13.04 (1.16)	13.04 (1.16)
Main noun frequency ^a	75.14 (112.66)	69.64 (97.03)
Final word frequency ^a	73.04 (110.68)	77.53 (123.02)
Main noun single-word dominance score ^b	0.75 (0.16)	
Sentence comprehensibility ^c	8.2 (1.04)	
Sentence level subordinate-meaning suitability rating ^d	7.9 (1.08)	

^aLemma frequency per million (CELEX lexical database, Baayen et al., 1995).

^bVolunteers in a pilot word association study, who did not participate in the present experiment, were presented with the ambiguous main nouns in isolation, and asked to respond with the first word that came to mind. After providing their association response, they selected their intended meaning of the ambiguous main noun from alternatives. The score provided here indicates the proportion of word association responses that matched the purported dominant meaning of the word.

^cPilot participants rated whether the sentences made sense on a scale from 1–10 (with 1 indicating "makes no sense at all", and 10 indicating "makes perfect sense", see Vitello, 2014; and Vitello et al., 2014, for details).

^dPilot participants were explicitly asked which of two possible meanings (e.g. "the noise a dog makes" or "the outer covering of a tree") made more sense in the given context for each sentence (see Vitello, 2014; and Vitello et al., 2014). The score provided here indicates the preference given to the subordinate meaning on a scale from 1–10 (with 1 indicated "subordinate meaning makes no sense at all", and 10 indicating "subordinate meaning makes perfect sense").

Therefore, the critical window for sentence reinterpretation began at the final word in the sentence.

Extensive stimulus piloting was conducted to confirm the dominance bias of key nouns used in the sentences (main noun single-word dominance score, see Table 2). In addition, piloting confirmed that when the whole sentence context including the disambiguating final word was presented, key noun meaning preference favoured the subordinate meaning (sentence-level subordinate meaning suitability rating, see Table 2). Piloting therefore confirmed that the sentences in the Ambiguous condition generally led comprehenders down a metaphorical "garden-path", with initial dominant meaning selection, followed by context favouring the subordinate meaning that necessitated sentence reinterpretation to achieve successful comprehension. For details about the piloting procedures please see the Supplementary Materials associated with Blott et al. (2021, <https://osf.io/hn3bu>). A list of all stimuli used in the present study can be found at <https://osf.io/m87vg/>.

For each Ambiguous sentence, an Unambiguous control version of the sentence was created by replacing the key noun with an unambiguous noun, and substituting a new final word that provided context consistent with the unambiguous word meaning (see Unambiguous example in Table 1). Key nouns and final words in Ambiguous and Unambiguous sentence versions were matched on syllable length and lemma frequency (stimulus characteristics are summarised in Table 2). The focus in these experiments was the process of post-sentence reinterpretation rather than within-sentence meaning selection, and the choice to construct matched Ambiguous and Unambiguous sentence pairs using the same sentence frame with different key nouns and final words was designed to create a balance between matching stimuli as closely as possible while reducing the risk of confounds; for example, limiting surface familiarity between stimulus pairs to avoid practice effects, and removing ambiguous key nouns from Unambiguous sentences to avoid inadvertent disambiguation.

In Experiment 1, a subset of 63 matched pairs of Ambiguous and Unambiguous sentences were selected at random from the full stimulus set, and were used as coherent experimental sentences. The remaining 27 sentence pairs from the full stimulus set were modified for use as semantically anomalous filler sentences, in which meaning coherence was violated; thus Anomalous filler sentences contained either an ambiguous (e.g. "scoop") or an unambiguous (e.g. "quest") key noun (see Anomalous conditions in Table 1). In the Anomalous filler sentences, the final word was altered so that the context it provided was always incompatible

with key noun meaning; for Anomalous sentences with ambiguous key nouns the final word was incompatible with both the dominant and the subordinate meaning. Key nouns were not duplicated across conditions.

All stimuli were prepared in three presentation formats: Spoken, Written, and RSVP. Auditory sentences were recorded by a female speaker of British English. Visual items were presented in black font on a white background.

Design

In this experiment, Ambiguous and Unambiguous, and Anomalous conditions were presented in each of the three formats; Spoken, Written and RSVP. Each participant completed a total of 180 experimental trials, divided equally between the three presentation formats. Of the 60 trials allocated to each format, 42 trials (70%) were coherent sentences (21 Ambiguous and 21 Unambiguous) and 18 trials (30%) were semantically anomalous fillers (9 with ambiguous and 9 with unambiguous key nouns). Stimuli were presented to each participant in three experimental blocks. Each block included two 10-trial miniblocks for each of the three presentation formats. Within each format-specific miniblock, 70% of trials were coherent (a combination of Ambiguous and Unambiguous trials), and 30% were anomalous. All participants encountered the format-specific miniblocks in the same pseudorandomised order but saw or heard different stimuli in these miniblocks. For individual participants, the matched coherent Ambiguous and Unambiguous sentence pairs were presented in the same format, but in different experimental blocks. The assignment of sentence pairs to the different presentation formats was counterbalanced across participants. This design ensured that, across participants, each ambiguous key noun appeared in all three presentation formats, but was presented only once, in a single format, to each individual participant. Only the coherent Ambiguous and Unambiguous conditions were included in the analyses.

Procedure

The experiment was run using the experiment platform Gorilla (www.gorilla.sc/about; Cauldron Inc.; Anwyl-Irvine et al., 2019). Participants completed the study online on their own computers using headphones. Before the main task, participants completed a demographic questionnaire on their age, sex, education and language background.

In the Meaning Coherence Judgement task, participants were instructed to listen to or read each sentence carefully for comprehension, and to decide as quickly and as accurately as possible whether each sentence

made sense or not. A visual symbol (an ear or eye image) was shown on the screen at the start of each miniblock to cue participants to expect either spoken or written stimuli. No additional cues were given as to the format of visual presentation (i.e. Written vs RSVP). After presentation of a sentence, two response buttons appeared (“Sense” and “Nonsense”) for participants to select with their mouse. The position of these response buttons on the screen was counterbalanced across participants.

In the Spoken condition, duration of sentence stimuli ranged between 2.8–4.0 s ($M_{\text{Duration}} = 3.5$ s). The presentation software was unable to support the automatic playing of sound files, so participants clicked a button on the screen with the mouse to begin playing each file manually. Response buttons appeared immediately after the end of the audio file. For the Written condition, whole sentences were presented on the screen for the mean duration of all audio files (3500 ms) plus an extra 1000 ms; response buttons appeared at the end of this 4500 ms interval. The presentation duration was designed to allow sufficient time to read and resolve Ambiguous stimuli while the sentence was still visible on the screen. For the RSVP conditions, each word in the sentence was presented consecutively in the centre of the screen for a duration of 250 ms; the end of the sentence was indicated by a full stop at the end of the final word. Response buttons appeared 250 ms after presentation of the final word. Trials were separated by a fixation cross presented in the centre of the screen for 2500 ms.

Before the experimental trials, participants practised the Meaning Coherence Judgement task on four coherent Unambiguous and two Anomalous Unambiguous sentences (two sentences per format) which were not repeated in the main experiment. Immediate feedback on response accuracy was given for the practice trials, but not during the main experiment. Participants took an enforced 30 s break between each experimental block, and the overall duration of the Meaning Coherence Judgement task was approximately 25–30 min.

Data analysis

Analyses were conducted in RStudio (v. 3.4.2; RStudio Team, 2015), using mixed effects modelling (Baayen et al., 2008; Barr et al., 2013). Analyses were conducted on the coherent Ambiguous and Unambiguous conditions only. Response times were log-transformed for all analyses. Trials in which response times were below 250 ms were removed from the analyses under the assumption that these were accidental button presses. Separate analyses were conducted for response accuracy and response time on correct trials, using the

glmer() function of the lme4 package (v.1.1.17; Bates et al., 2015) with the bobyqa optimiser, and the lmer() function from the same package, respectively. Hypothesis tests were based on model comparisons by likelihood ratio tests, and Chi-squared and p -values (with α set at .05, unless stated otherwise) from these tests are reported. All models included fixed effects for Ambiguity (deviation-coded variables: Ambiguous $-1/2$, Unambiguous $1/2$), Format (Format.code1 with Spoken deviation-coded as $1/3$, Written as $1/3$, RSVP as $-2/3$, and Format.code2 with Spoken deviation-coded as $1/2$, Written as $-1/2$, and RSVP as 0), and their interactions. Models with maximal random effects structures (a random intercept by items, and a random intercept and slope for Ambiguity, Format, and their interaction by subjects) were fitted where possible (Barr et al., 2013). In case of convergence issues, we went through the following steps iteratively until the model converged: (1) remove correlations between the random effects by subjects, (2) remove random intercept by subjects, (3) remove the random slope that explains the least variance in the maximal model. We treated the Format codes and Ambiguity x Format interaction codes as pairs, and our reports of main effects of Format and an Ambiguity x Format interaction are based on removing the pair of relevant codes, unless specified.

Results

Accuracy

The maximal model that would converge contained fixed effects for Ambiguity, Format, and their interaction, as well as a random intercept by items and a random intercept and random slopes for Ambiguity and the Ambiguity x Format interaction by subjects, without accounting for correlations between the random effects. Detailed results from model comparisons can be found in supplemental materials. There was a significant main effect of Ambiguity, with Ambiguous sentences associated with lower accuracy ($M_{\text{Total}} = 0.77$,

$SD_{\text{Total}} = 0.42$) than their Unambiguous counterparts ($M_{\text{Total}} = 0.94$, $SD_{\text{Total}} = 0.23$; see Figure 1, and Table 3; $\beta = 1.75$, $SE = 0.07$, $z = 24.58$; Model comparison: $\chi^2(1) = 216.45$, $p < .001$). There was also a significant main effect of Format, with the RSVP format associated with the lowest accuracy rates (see Figure 1, and Table 3; Format.code1 (Spoken and Written > RSVP): $\beta = -0.35$, $SE = 0.07$, $z = -5.44$; Format.code2 (Spoken > Written): $\beta = -0.94$, $SE = 0.08$, $z = -0.53$; Model comparison: $\chi^2(2) = 28.24$, $p = .002$). The Ambiguity x Format interaction, however, was not statistically significant ($p = .833$). Since accuracy of Meaning Coherence judgements for Ambiguous sentences reflects successful retrieval and integration of the subordinate meaning of the key noun, the lack of a significant Ambiguity x Format interaction suggests that presentation format did not significantly influence participants' ability to resolve ambiguity successfully.

Response time

The maximal model that converged contained fixed effects for Ambiguity, Format, and their interaction, as well as a random intercept by items and a random intercept and random slopes for Format by subjects, including correlations between random effects. Detailed results from model comparisons can be found in supplemental materials. Response times on correct trials in the Meaning Coherence Judgement task were significantly longer when participants needed to resolve an ambiguity to arrive at the correct interpretation ($M_{\text{Total}} = 1213.77$, $SD_{\text{Total}} = 928.91$) than when the sentence was unambiguous ($M_{\text{Total}} = 993.3$, $SD_{\text{Total}} = 737.66$; see Figure 1 and Table 3; $\beta = -0.08$, $SE = 0.004$, $t = -19.89$; Model comparison: $\chi^2(1) = 388.5$, $p < .001$). There was also a significant main effect of Format on response times (see Figure 1 and Table 3; Format.code1 (Spoken and Written > RSVP): $\beta = -0.01$, $SE = 0.01$, $t = -1.03$; Format.code2 (Spoken > Written): $\beta = 0.13$, $SE = 0.01$, $t = 16.5$; Model comparison: $\chi^2(2) = 135.57$, $p < .001$). Importantly, these main effects were qualified by a significant Ambiguity x Format interaction, suggesting that the magnitude of ambiguity-related costs to response times depended on the format in which sentences had been presented (see Figure 1 and Table 3; Ambiguity x Format.code1 (Spoken and Written > RSVP): $\beta = -0.04$, $SE = 0.01$, $t = -4.35$; Ambiguity x Format.code2 (Spoken > Written): $\beta = -0.06$, $SE = 0.01$, $t = -6.76$; Model comparison: $\chi^2(2) = 64.45$, $p < .001$).

We decomposed this interaction by running separate mixed models with Ambiguity as a fixed factor, a random intercept by items and random intercept and random Ambiguity slope by subjects for each format. Detailed results from model comparisons can be found in

Table 3. Summary of descriptive statistics for accuracy and response times in the Meaning Coherence Judgement task for each of the presentation formats (means with standard deviations in parentheses).

	Spoken	Written	RSVP
Accuracy			
Ambiguous	0.8 (0.13)	0.8 (0.14)	0.76 (0.14)
Unambiguous	0.95 (0.06)	0.95 (0.07)	0.93 (0.09)
Mean difference	0.15	0.15	0.18
RT			
Ambiguous	1413.5 (435.1)	981.4 (351)	1223.7 (373.7)
Unambiguous	1122 (336.9)	887.9 (299.6)	961.3 (347)
Mean difference	291.5	93.5	262.4

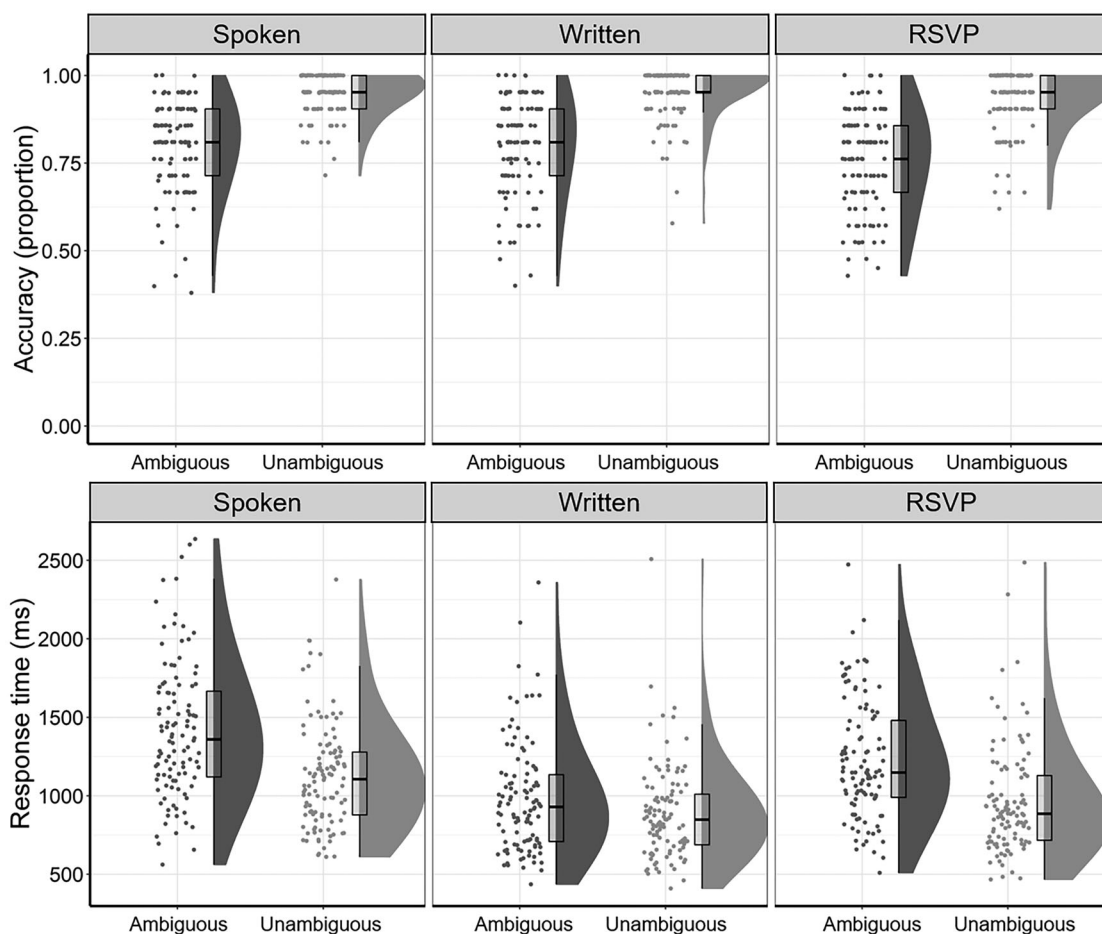


Figure 1. Accuracy rates and response times in the Meaning Coherence Judgement task. Boxplots show median and interquartile range of accuracy scores (top panel) and response times (bottom panel) for the Ambiguous and Unambiguous condition in each presentation format.

supplemental materials. Although there were statistically significant ambiguity-related response time costs in the Spoken format ($\beta = -0.1$, $SE = 0.01$, $t = -13.91$; Model comparison: $\chi^2(1) = 111.25$, $p < .001$), in the Written format ($\beta = -0.03$, $SE = 0.01$, $t = -4.8$; Model comparison: $\chi^2(1) = 22.99$, $p < .001$; maximal model was an intercept-only model), and in the RSVP format ($\beta = -0.1$, $SE = 0.01$, $t = -13.76$; Model comparison: $\chi^2(1) = 110.2$, $p < .001$), the ambiguity costs to response times were exaggerated relative to the Written condition in the Spoken and RSVP format.

Discussion

Experiment 1 provided a clear demonstration of the effects of ambiguity and presentation format on higher-order language comprehension processes. Consistent with previous studies, our results confirmed that lexical-semantic ambiguity impedes successful sentence comprehension: the presence of ambiguity significantly reduced the likelihood of accurate

comprehension, even in the context of an explicit comprehension task, and significantly increased response times relative to an unambiguous baseline. Additionally, the RSVP format was associated with significantly reduced comprehension accuracy compared to the Spoken and Written formats, irrespective of the presence or absence of ambiguity. This finding suggests that the relatively unnatural word-by-word presentation format reduced participants' general ability to build up coherent sentence-level meaning representations (see also e.g. Masson, 1983; Schotter et al., 2014). We were particularly interested in the modulation of ambiguity processing costs by presentation format. Although format did not affect the cost of ambiguity to comprehension accuracy, we did find that format significantly influenced the temporal costs of disambiguation, resulting in exaggerated "garden-path" effects to response times for both the RSVP and Spoken formats (mean effect sizes of about 260 and 290 ms), compared to the whole-sentence Written format (mean effect size of about 90 ms). The present data do not suggest evidence for different

speed-accuracy tradeoff patterns in the different formats. Although participants tended to respond faster in the Written format compared to the other formats, the Written format was not also associated with a greater likelihood for errors. Similarly, although the ambiguity effect on response times was largest for the Spoken and RSVP formats, the ambiguity effect on accuracy was at least as big as in the Written format. Taken together, these findings suggest that, in contrast to the whole-sentence Written format, successful disambiguation in the RSVP and Spoken formats was achieved at the expense of increased processing time. We will return to potential explanations of this finding in the General Discussion.

Experiment 2: neural processing costs of ambiguity resolution

Complementing the behavioural results from the explicit comprehension task in Experiment 1, Experiment 2 used functional MRI during implicit language processing to investigate whether presentation format modulated neural activity within cortical language regions that have previously been shown to be involved in the processing of ambiguity and reinterpretation. We expected to replicate previous findings of ambiguity-related increases in haemodynamic responses in typical language regions including left IFG, compared to the control sentences (e.g. Mason & Just, 2007; Rodd et al., 2005; Zemleni et al., 2007). In Experiment 2, the key interest lay in determining whether the magnitude of ambiguity-related effects in this region was influenced by how sentences were presented. We predicted that, in line with the differential effects of presentation format on the neural activation costs of syntactic complexity that have been previously reported (Lee & Newman, 2009), we would observe significant differences between the presentation formats in terms of ambiguity-related neural activity.

Method

Participants

Data from 17 monolingual native speakers of English (12 female, aged 20–44, $M_{\text{Age}} = 24.8 \pm 5.4$ years) were included in the analyses. An additional 5 participants were excluded from analyses due to withdrawal of consent, or technical difficulties with data collection. Participants were recruited from the UCL SONA participant pool, and had on average spent 6 years ($SD = 1.4$, range: 2–8 years) in formal education settings after age 16. All participants were self-reported right-handers, had no hearing or uncorrected visual impairments, nor

any history of neurological illness or head injury. UCL ethical approval for this study was obtained via the Birkbeck-UCL Centre for Neuroimaging (BUCNI) departmental ethics chair. Written informed consent was obtained from all participants, and they were financially compensated for participation at standard UCL rates.

Materials

Sentence stimuli for Experiment 2 consisted of the full set of 90 matched pairs of Ambiguous and Unambiguous sentences (see Table 2). No semantically anomalous sentences were used in Experiment 2. Stimuli for the three presentation formats, Spoken, Written and RSVP, were created for the Ambiguous and Unambiguous versions of each sentence. For the Spoken format, each sentence was recorded digitally by a female speaker of British English. For the Written format, to enable consistent stimulus presentation with the software used, digital images were created of each sentence written as 2–3 lines of text in black font on a white background. Similarly, for the RSVP format, separate digital images of each word in a sentence, written in black font on a white background, were created.

In order to enable identification of brain regions that demonstrated preferential responses to intelligible language, Experiment 2 also included low-level unintelligible baseline conditions for each presentation format, created by modifying the original sentence stimuli. For the Spoken condition, Unintelligible stimuli were created by spectral rotation (around a 2kHz frequency) of the original sentences, following the signal processing procedure described by Scott and colleagues (2000; see also Blesser, 1972). This technique preserves the acoustic complexity of speech stimuli but renders them unintelligible. For Written and RSVP conditions, Unintelligible stimuli were created by substituting each letter in the original sentences with a corresponding false font symbol. The false font was adapted from an obsolete near-Eastern alphabet whose characters had a letter-like structure but minimal direct resemblance to Roman letters (Melchert, 2004), meaning that false font characters had limited links to phonological representations for our participants (Warren, 2013). As with the intelligible conditions, Unintelligible stimuli for the Written format were digital images of whole-sentence stimuli written as 2–3 lines of false-font “text”, while stimuli for the RSVP format were separate digital images of each “word” in a sentence; stimuli were black font on a white background.

fMRI data acquisition

Whole-brain functional images were collected on a Siemens Avanto 1.5-T MR scanner with a 32-channel

head coil, using a continuous-acquisition method. Echo-planar image volumes (232–251 volumes per participant) were acquired over three runs of approximately 12 min each, using a gradient-echo EPI sequence (TR = 3000 ms, TE = 50 ms, isometric 3 mm voxels). In addition, anatomical images were acquired using a T1-weighted MPRAGE sequence (TR = 2730 ms, TE = 3.57 ms, isometric 1 mm voxels, duration 5.5 min).

Design

In this experiment, Ambiguous, Unambiguous, and Unintelligible conditions were presented in each of the three formats; Spoken, Written and RSVP. There were no Anomalous filler trials in Experiment 2. Each participant completed a total of 234 experimental trials, divided equally between the three presentation formats. Of the 78 trials allocated to each format, 30 trials were Ambiguous sentences, 30 were Unambiguous sentences, and 18 were Unintelligible stimuli. Stimuli were presented to each participant in three experimental runs. Each run included two 13-trial miniblocks for each of the three presentation formats. Each format-specific miniblock included 5 trials each of the Ambiguous and Unambiguous conditions, and 3 Unintelligible trials. Three pseudorandomised miniblock orders were assigned randomly across runs for each participant. Within each miniblock, Ambiguous, Unambiguous and Unintelligible trials were presented in a pseudorandomised order. For individual participants, the Ambiguous, Unambiguous and Unintelligible versions of a given sentence were presented in the same format, but in different experimental runs. The assignment of corresponding Ambiguous, Unambiguous and Unintelligible sentence versions to the different presentation formats was counterbalanced across participants. As in Experiment 1, this design ensured that, across participants, each ambiguous key noun appeared in all three presentation formats, but was presented only once, in a single format, to each individual participant.

Procedure

Stimuli were presented using MATLAB (Mathworks Inc., Natick, MA, USA) and the Cogent 2000 toolbox (www.vislab.ucl.ac.uk/cogent/index.html). Auditory stimuli were presented binaurally through MRI-compatible insert earphones (Sensimetrics, Malden, MA, USA, Model S-14), which provided a manufacturer-evaluated scanner noise attenuation level of 20–40 dB. Visual stimuli were projected onto a screen placed in front of the bore of the MRI scanner and viewed by an angled mirror above the participant's head. Sound volume levels and stimulus readability were checked prior to the first functional run using spoken and written

stimuli which were not repeated in the experiment proper. Participants also heard and saw examples of the unintelligible stimuli before commencing the experiment.

Participants were instructed to listen to or read each sentence carefully for comprehension during scanning. An explicit task was avoided in Experiment 2 in order to avoid confounding neural activation related to ambiguity resolution with activation related solely to the non-specific demands of an overt task (Wright et al., 2011). Participants were informed prior to scanning that some trials would consist of nonsense sounds or letters, and were instructed to pay attention to these trials even though the stimuli would be incomprehensible. For the Unintelligible Written condition, participants were given the additional instruction to move their eyes across the line of false font text from left to right as if reading English.

As in Experiment 1, a visual symbol (an ear or eye image) shown on the screen at the start of each miniblock cued participants to expect either spoken or written stimuli, with no additional cues given to distinguish between visual presentation formats. During the Spoken conditions, duration of sentence stimuli ranged between 2.4–4.0 s ($M_{\text{Duration}} = 3.1$ s). For the written conditions, whole-sentence stimuli were displayed on the screen for the same duration as the corresponding auditory sentence stimulus plus an additional 1500 ms; this presentation duration was designed to allow sufficient time to read and resolve Ambiguous stimuli while the sentence was still visible on the screen. For the RSVP conditions, each word in the sentence was presented consecutively in the centre of the screen for a duration of 250 ms; the end of the sentence was indicated by a full stop at the end of the final word. Thus, the presentation parameters for all three sentence formats were as close as possible to Experiment 1.

Trials were separated by a fixation cross presented in the centre of the screen. Fixation cross display time was 2500 ms plus a randomly-jittered interval of 1000–3000 ms, so that the overall inter-trial interval varied between 3500–5500 ms. Participants took short (30–60 s) breaks between each experimental run. Anatomical images were obtained after the completion of the three functional runs.

As fMRI scanning did not involve an explicit comprehension task, two computerised behavioural tests were conducted immediately after scanning to obtain explicit measures of stimulus engagement during scanning (Recognition Memory test) and disambiguation ability (Explicit Disambiguation test). These tests were conducted outside the scanner approximately 5–10 min after completion of the scanning session, with each

test lasting approximately 15 min; participants were not warned prior to scanning that these tests would be conducted. Both tests were delivered outside the scanner in a quiet testing environment on a laptop using the Cogent toolbox running in MATLAB. The Recognition Memory test assessed recall of experimental sentence stimuli, and served as a check that participants had paid attention to the stimuli presented during scanning despite the absence of an overt task. A subset of 72 Ambiguous fMRI stimuli (sampled evenly across the experiment; 3 from each miniblock within each run) plus 72 foil ambiguous sentences with the same subordinate-biased late-disambiguating structure (not presented during scanning) were presented in randomised order to participants, who were instructed to read each sentence and to decide whether they had seen or heard it during scanning. Unambiguous sentence stimuli were not included in the Recognition Memory Test to limit the length of the test. All fMRI and foil sentences were presented in written whole-sentence format, irrespective of the presentation format of the Ambiguous stimuli during scanning.

Following the Recognition Memory test, the Explicit Disambiguation test was used to confirm that individual participants could successfully disambiguate the specific sentences used in the fMRI experiment. This test was also designed to generate approximate measures of whole-sentence reading time for the Written condition, for use in fMRI analyses (see *Image Analysis*). Stimuli were taken from the fMRI experiment and included all 90 Ambiguous sentences, as well as 30 randomly selected Unambiguous sentences included to provide a control condition without increasing the length of the task too much. All stimuli were presented in randomised order in written whole-sentence format, irrespective of their presentation format during scanning. Participants were instructed to press a button as soon as they had read through the whole sentence once; button press latency provided a measure of whole-sentence reading time. After reading each sentence, participants were then asked to select the correct meaning of the key noun within the sentence context. Three response options were presented for key-noun meaning in each Ambiguous sentence: the subordinate (correct) meaning, the dominant (incorrect) meaning, and a third “other” option. For Unambiguous sentences, key-noun meaning response options were the correct meaning, an incorrect unrelated meaning, and “other”. Participants were instructed to select the meaning they had attributed to the word after understanding the whole sentence in the scanner, or, if unable to recall the sentence, to choose the meaning most appropriate to the sentence context. The position of the correct meaning was randomised across trials.

Data analysis

Behavioural tests. Analyses of behavioural data were conducted in RStudio (v. 3.4.2; RStudio Team, 2015), by means of mixed effects models with random effects for subjects and items (Baayen et al., 2008; Barr et al., 2013). Following procedures for confirmatory hypothesis tests outlined in Barr and colleagues (2013), model complexity was reduced only when there were convergence issues, based on the same iterative steps as detailed in the Method section for Experiment 1.

Analysis models for the Recognition Memory test were constructed to investigate the effects of sentence presentation format and scanning run on recognition memory accuracy; logit mixed effects models were fitted using the `glmer()` function with the `bobyqa` optimiser from the `lme4` package (v. 1.1.17; Bates et al., 2015).

Accuracy scores and reading times for correct trials in the Explicit Disambiguation post-test were analysed to test for behavioural effects of Ambiguity. Reading times were calculated as the latency from stimulus appearance to button press and log-transformed prior to entry into linear mixed effects models, which were fitted with the `lmer()` function from the `lme4` package (v. 1.1.17; Bates et al., 2015). We also explored whether the sentence presentation format during scanning affected disambiguation accuracy in the post-test. Logit mixed effects models were fitted to data from the Ambiguous sentences only with the `glmer()` function within the `lme4` package (v. 1.1.17; Bates et al., 2015).

fMRI data. The functional images from the fMRI experiment were preprocessed and analysed using SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK). Preprocessing steps included within-subject realignment, within-subject co-registration of the T1-weighted structural image to the mean EPI image, and spatial normalisation of the EPI images to standard MNI space by the unified segmentation method (as described in Ashburner & Friston, 2005). The data were spatially smoothed using an 8 mm FWHM Gaussian kernel. For one participant, a single run was excluded from the fMRI analyses due to excessive head motion.

Statistical analyses were performed using the general linear model and Gaussian random field theory as implemented in SPM12 (Friston et al., 1994). Each of the three functional runs was modelled separately. The design matrices included onsets and durations for the miniblock presentation format cues (the ear or eye symbols), movement parameters, and temporal and dispersion derivatives. Rest served as an implicit baseline. A high-pass filter of 128 s was applied to remove low-

frequency noise. Two separate models were used to analyse the data in each presentation format: Intelligibility (intelligible > unintelligible), and Ambiguity (Ambiguous > Unambiguous) contrast models.

Intelligibility contrast models

An initial model was used to investigate activation related to within-sentence processing of intelligible language in each presentation format at the whole-brain level. For each participant, sentence onset and duration for all intelligible stimuli (Ambiguous plus Unambiguous) and for the unintelligible stimuli were modelled using separate regressors for each presentation format. Ambiguous and Unambiguous stimuli were combined into a single Intelligible condition for each format because stimuli were designed so that sentence processing did not differ between these stimuli until the sentence-terminal word. Individual-subject contrast images from the comparison of Intelligible > Unintelligible conditions were entered into three separate second-level one-sample *t*-test models to determine format-specific within-sentence intelligibility effects at the group level. This analysis was used to identify region preferentially responsive to intelligible language, so that activation within language-responsive cortex could be identified in the subsequent whole-brain analyses of ambiguity effects.

Ambiguity contrast models

The initial model was elaborated in order to investigate post-sentence activation related to ambiguity resolution in each presentation format at the whole-brain level. For each participant, in addition to the modelling of format-specific sentence onset and duration for the combined Intelligible (i.e. Ambiguous and Unambiguous) and the Unintelligible conditions, sentence offsets for the Ambiguous and the Unambiguous sentences were added to the model as separate sets of events for each presentation format. Sentence offset was modelled explicitly because this was the point at which processing of Ambiguous and Unambiguous sentences was designed to diverge; in Ambiguous sentences, the offset of the disambiguating sentence-terminal word was assumed to trigger semantic mismatch detection and reinterpretation processes (Rodd et al., 2012; Twilley & Dixon, 2000). The addition of sentence-offset modelling inevitably resulted in a degree of temporal overlap between within-sentence and post-sentence neural activity. However, this model did enable the effects of ambiguity resolution on post-sentence activity to be at least partly separated from general within-sentence processing of intelligible language. In the

Spoken and RSVP conditions, sentence offset corresponded to the end point of stimulus presentation. In the Written condition, the reading times obtained for each stimulus in the Disambiguation post-test were used as estimations of individual sentence offset times for each participant (for a small number of outlier reading times that differed by more than 1500 ms from the duration of the corresponding sentence's audio file in the Spoken condition, the audio file duration was substituted as a proxy offset time). Individual-subject contrast images from first-level analyses were entered into second-level models to determine ambiguity effects, and interactions between ambiguity and format at the group-level.

Region-of-interest analyses

In addition to whole-brain analyses, the interaction between ambiguity and presentation format effects was explored using regions-of-interest (ROI) analyses. ROI location was determined independently of the current dataset using activation peaks from previous studies that have contrasted sentence-offset activity for subordinate-constrained ambiguous sentences and unambiguous control sentences presented in auditory format (Rodd et al., 2013), and presented phrase-by-phrase in visual format (Zempleni et al., 2007). Peaks from these two studies were used to create ROIs in the present study only if they were also present in our intelligible language contrasts. This procedure ensured that ROIs in the current study were located in regions which were likely to be responsive to meaningful language stimuli in our sample (based on their presence in the group-level intelligible > unintelligible contrast), and have also previously been found to be responsive to manipulations of semantic ambiguity (although see Poldrack, 2006, for the difficulties with reverse inferences from anatomical locations to cognitive function in the absence of participant-specific functional localisers). Only two peaks fulfilled these selection criteria, both located within left IFG: a more anterior and ventral peak located within BA 47 (MNI coordinates $[-50, 30, -2]$), and more posterior and dorsal peak located within BA 45 ($[-50, 27, 11]$). Once activation peaks fulfilling these criteria had been identified, the MarsBar toolbox for SPM 12 (Brett, Anton, Valabregue, & Poline, 2002) was used to create spherical 8mm-radius ROIs centred on each of these peaks, and individual-participant mean differences in sentence-offset responses to Ambiguous versus Unambiguous stimuli were extracted from each ROI for each presentation format. The resulting difference scores were entered into a one-way ANOVA (JASP v.0.9.2; JASP Team, 2018) to identify effects of presentation

format on the difference scores; significant effects were examined further by means of paired-samples t-tests.

Results

Behavioural results

The mean proportion of correct responses for Ambiguous sentences in the Recognition Memory test indicated adequate levels of attention during scanning ($M_{\text{Total}} = 0.66$, $SD_{\text{Total}} = 0.2$, range 0.36–0.94). Detailed results from model comparisons can be found in supplemental materials. Comparisons between scanning runs demonstrated no significant effect of the interval between scanning and testing on Recognition Memory (Run 2 and Run 3 > Run 1: $\beta = -0.15$, $SE = 0.17$, $z = -0.92$; Run 2 > Run 3: $\beta = 0.05$, $SE = 0.17$, $z = 0.31$; Model comparison: $\chi^2(2) = 0.87$, $p = .644$, maximal model did not account for correlations between random intercept and slopes by subjects). One-

sample t-tests demonstrated that participants' Recognition Memory accuracy for sentences in the Spoken ($M = 0.57$, $SD = 0.24$), Written ($M = 0.74$, $SD = 0.18$) and RSVP ($M = 0.67$, $SD = 0.21$) conditions were significantly greater than chance levels of 0.5 (Spoken: $t(407) = 3$, $p = .003$; Written: $t(407) = 10.9$, $p < .001$; RSVP: $t(407) = 7.13$, $p < .001$; all compared against a Bonferroni-corrected α -level of .017), as can also be seen in Figure 2. There was, however, a significant main effect of in-scanner presentation format on participants' ability to recall Ambiguous scanning stimuli during the Recognition Memory test (see Figure 2), suggesting that memory for the Spoken condition was comparatively poor (Spoken and Written > RSVP: $\beta = 0.04$, $SE = 0.14$, $z = 0.28$; Spoken > Written: $\beta = -0.9$, $SE = 0.17$, $z = -5.39$; Model comparison: $\chi^2(2) = 29.94$, $p < .001$, maximal model was an intercept-only model).

The mean proportion of correct responses to Ambiguous stimuli in the post-scan Explicit Disambiguation

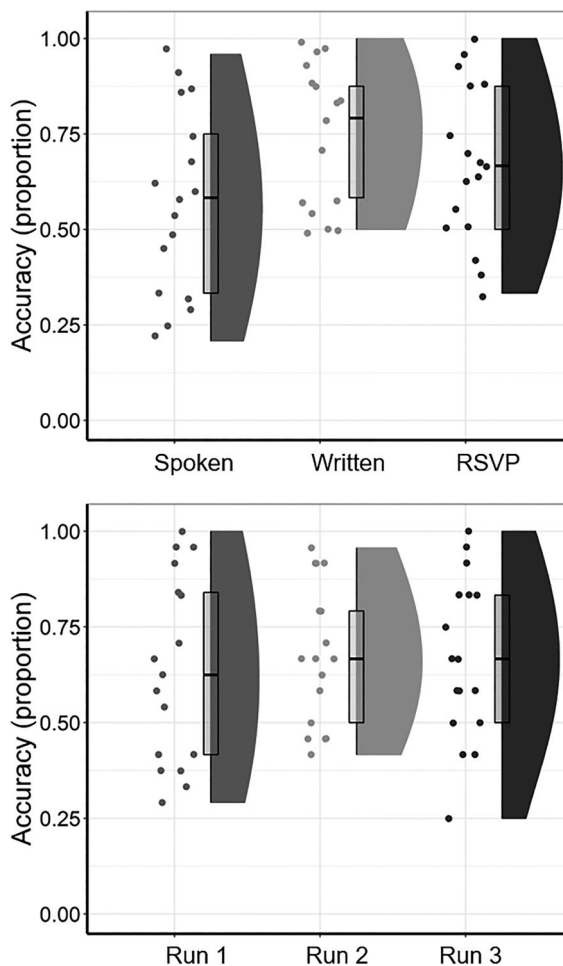


Figure 2. Accuracy rates in the post-scan Recognition Memory test. Boxplots show median and interquartile range of accuracy scores, by scanning run (top panel) and presentation format (bottom panel).

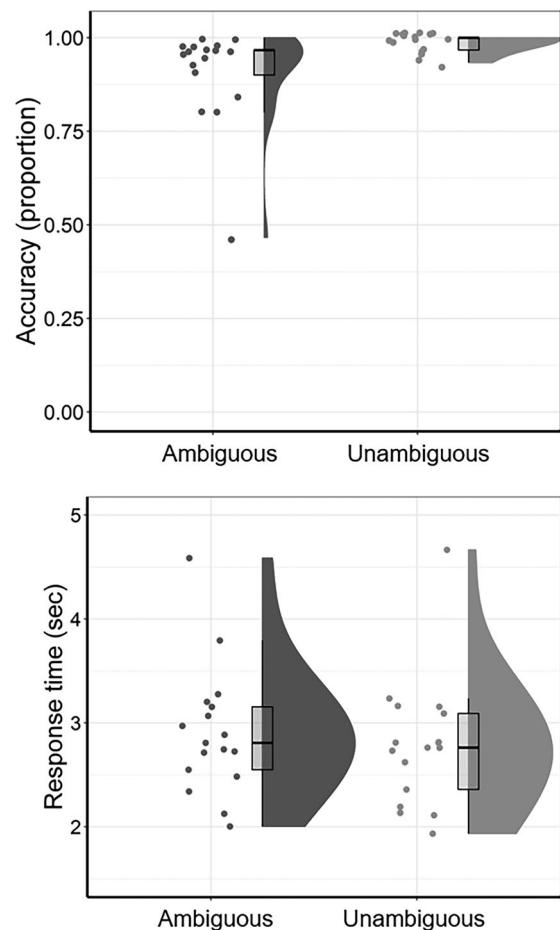


Figure 3. Accuracy rates and response times in the post-scan Explicit Disambiguation task. Boxplots show median and interquartile range of accuracy scores (top panel) and response times (bottom panel) for the Ambiguous and Unambiguous condition.

test demonstrated that participants were generally able to successfully disambiguate the Ambiguous sentences used in the experiment ($M_{\text{Total}} = 0.91$, $SD_{\text{Total}} = 0.11$, range 0.56–1.00), though it should be noted that the Disambiguation task could not differentiate between successful immediate disambiguation at the time of first stimulus presentation during scanning, and delayed disambiguation at the time of second stimulus presentation during the post-scan task. As shown in Figure 3, significant effects of Ambiguity were observed on disambiguation accuracy ($\beta = 1.91$, $SE = 0.77$, $z = 2.47$; Model comparison: $\chi^2(1) = 7.72$, $p = .006$), and on sentence reading times of correct trials ($\beta = -0.02$, $SE = 0.01$, $t = -2.27$; Model comparison: $\chi^2(1) = 4.83$, $p = .028$). Detailed results from model comparisons can be found in supplemental materials. Ambiguous sentences were associated with significantly lower accuracy ($M_{\text{Acc}} = 0.91$, $SD_{\text{Acc}} = 0.28$) and longer reading times ($M_{\text{RT}} = 2.92\text{s}$, $SD_{\text{RT}} = 1.07\text{s}$) than Unambiguous sentences ($M_{\text{Acc}} = 0.99$, $SD_{\text{Acc}} = 0.11$; $M_{\text{RT}} = 2.77\text{s}$, $SD_{\text{RT}} = 1.21\text{s}$). No significant effects of scanning presentation format or run on disambiguation accuracy (Ambiguous trials only) were observed (Spoken and Written > RSVP: $\beta = 0.12$, $SE = 0.26$, $z = 0.46$; Spoken > Written: $\beta = -0.22$, $SE = 0.3$, $z = -0.72$; Model comparison: $\chi^2(2) = 0.15$, $p = .927$; Run 2 and Run 3 > Run 1: $\beta = -0.25$, $SE = 0.19$, $z = -1.29$; Run 2 > Run 3: $\beta = -0.002$, $SE = 0.24$, $z = 0.01$; Model comparison: $\chi^2(2) = 1.51$, $p = .471$, maximal models were intercept-only models, see supplemental

materials for detailed results from these model comparisons).

Whole-brain analyses

Panel 1 in Figure 4 shows the results of the basic contrast of intelligible > unintelligible stimuli. In the Spoken condition, haemodynamic responses to intelligible speech were contrasted with responses to spectrally rotated speech, which revealed activations in a fronto-temporal network of regions usually associated with speech comprehension (see Price, 2012, and Tyler & Marslen-Wilson, 2008 for reviews). In both written conditions, responses to meaningful written sentences were compared with responses to false font stimuli. Activations associated with the processing of meaningful visual language stimuli were observed in a fronto-temporal network, and therefore consistent with previous studies (e.g. Price, 2012; Tyler & Marslen-Wilson, 2008). Peaks from this contrast can be found in supplemental materials.

Panel 2 in Figure 4 illustrates the contrast of Ambiguous > Unambiguous sentences in each presentation format. Ambiguity effects were found in each format (peaks can be found in supplemental materials). The location of ambiguity-related activations was similar to those observed in previous studies of semantic ambiguity resolution in the auditory format (e.g. Rodd et al., 2005; Rodd et al., 2012), and in the visual format (e.g.

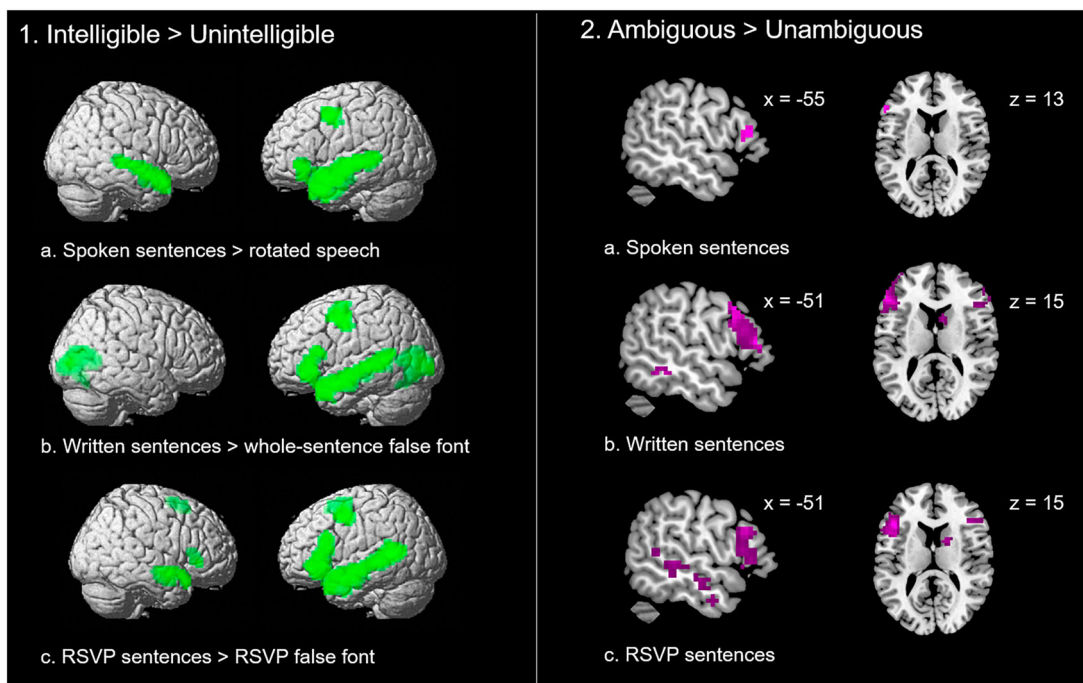


Figure 4. Results of the whole-brain analyses. Brain activation in the contrast of intelligible > unintelligible sentences (panel 1), and ambiguous > unambiguous sentences (panel 2) in each of the presentation formats.

Zempleni et al., 2007). Ambiguity effects were present in left frontal areas in all three presentation formats.

Region-of-interest results

The location of the two left IFG ROIs within BA 47 and BA 45 is illustrated on the left in panel 1 of Figure 5. A 3 (Format) \times 2 (Region) ANOVA comparing the ambiguity-related activation [Ambiguous > Unambiguous] in the three presentation formats across the two ROIs, revealed a significant main effect of Format, $F(2,32) = 4.23$, $p = .023$, $\eta^2 = 0.21$, on the ambiguity effects. Neither the main effect of Region, $F(1,16) = 0.02$, $p = .901$, $\eta^2 = 0.001$, nor the Format \times Region interaction, $F(2,32) = 2.32$, $p = .114$, $\eta^2 = 0.13$, were statistically significant.

Although the results of this ANOVA did not provide evidence of distinct ambiguity response profiles in the two IFG subregion ROIs, in view of the distinct language processing roles attributed to these regions previously

(see e.g. Binder et al., 2009; Thompson-Schill et al., 1997), it was of theoretical interest to confirm the effects of Format on ambiguity responses in each ROI separately. Therefore, a 1 \times 3 ANOVA was conducted for each ROI, comparing the difference in activation level in response to Ambiguous versus Unambiguous sentences, and was followed-up with individual t-tests. In the more anterior BA 47, there was a significant main effect of Format, $F(2,32) = 4.59$, $p = .018$, $\eta^2 = 0.22$, on the observed ambiguity effects (Ambiguous > Unambiguous contrast). One-sample t-tests demonstrated large ambiguity effects significantly greater than zero in the Written ($t(16) = 3.64$, $p = .002$, Cohen's $d = 0.88$), and the RSVP condition ($t(16) = 4.66$, $p < .001$, Cohen's $d = 1.13$), but a non-significant effect in the Spoken condition ($t(16) = 2.09$, $p = .053$, Cohen's $d = 0.51$). Post-hoc paired-samples t-tests comparing the ambiguity effects between the presentation formats revealed that ambiguity effects in the RSVP modality were significantly

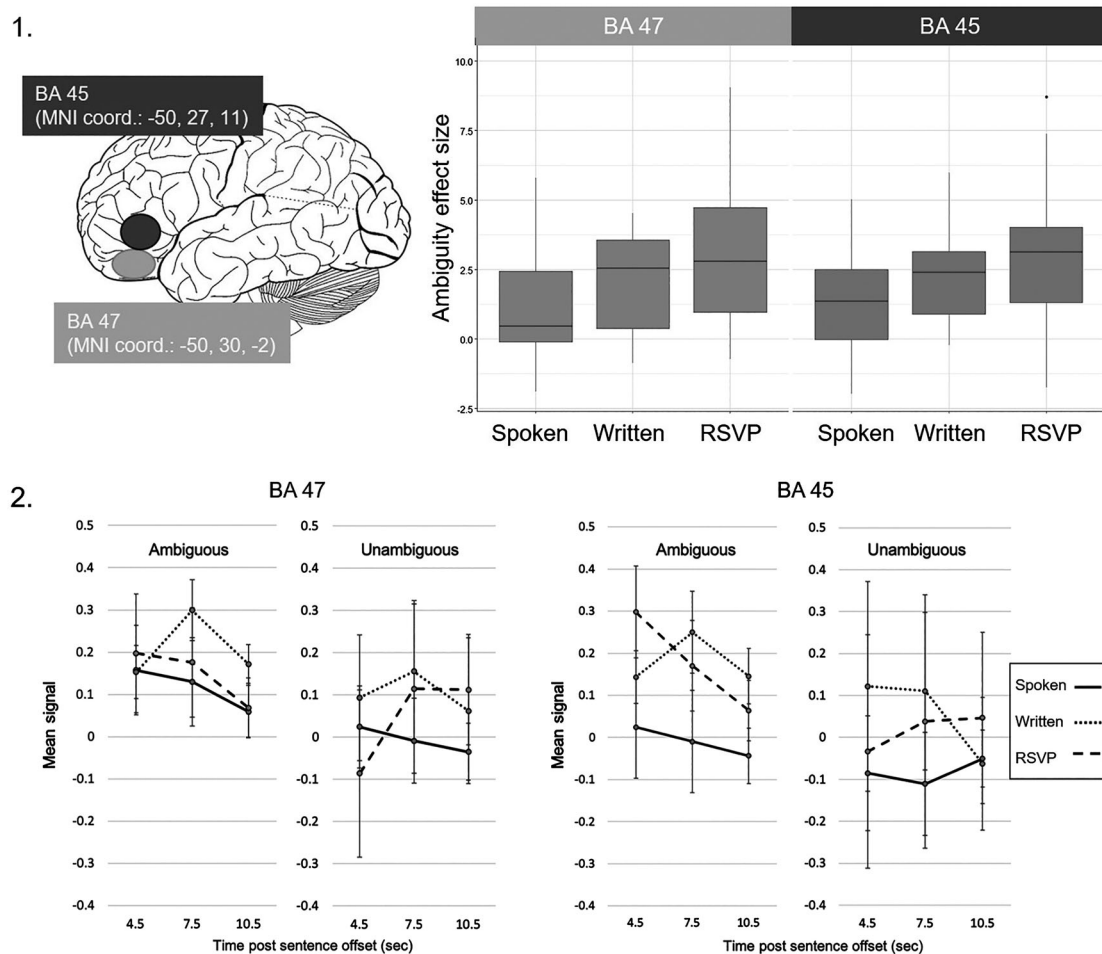


Figure 5. Results of the ROI analyses. Panel 1. Illustration of the approximate location of the BA 45 and BA 47 ROIs in the left IFG (left). Boxplots showing median and interquartile range of ambiguity effects (Ambiguous > Unambiguous contrast) in each of the presentation formats, and ROIs (right). Panel 2. Peri-stimulus time histograms of mean BOLD signal extracted from BA 47 (left) and BA 45 (right).

larger than those in the Spoken condition ($t(16) = 2.73, p = .015$, Cohen's $d = 0.66$), and those in the Written condition ($t(16) = 2.18, p = .044$, Cohen's $d = 0.53$; see Figure 5). The magnitude of ambiguity effects did not significantly differ between Spoken and Written conditions ($t(16) = -0.79, p = .441$, Cohen's $d = 0.19$). Inspection of peri-stimulus time histograms extracted from the BA47 region for each Ambiguous and Unambiguous condition suggested for the Written and RSVP conditions, the ambiguity effect represented a change from a low level of activity in this region in the Unambiguous conditions to a higher level of activity in the Ambiguous conditions (see panel 2 in Figure 5).

In the more posterior BA 45, the main effect of presentation format on the observed ambiguity effects was not statistically significant, $F(2,32) = 3.11, p = .058, \eta^2 = 0.16$. One-sample t-tests demonstrated ambiguity effects significantly greater than zero in all three presentation formats (Spoken: $t(16) = 2.84, p = .012$, Cohen's $d = 0.69$; Written: $t(16) = 5.78, p < .001$, Cohen's $d = 1.4$; RSVP: $t(16) = 4.34, p < .001$, Cohen's $d = 1.05$). Similar to BA 47, post-hoc pair-wise comparisons between presentation formats using paired-samples t-tests demonstrated significantly larger ambiguity effects in the RSVP condition than the Spoken condition, $t(16) = 3.06, p = .008$, Cohen's $d = 0.74$. However, the magnitude of ambiguity effects in the Written condition, though numerically intermediate between the effects observed in the other two conditions, did not differ significantly from either of them (Written vs. RSVP: $t(16) = -0.64, p = .532$, Cohen's $d = 0.16$); Written vs. Spoken: ($t(16) = 1.78, p = .094$, Cohen's $d = 0.43$). Inspection of peri-stimulus time histograms extracted from the BA45 region for each Ambiguous and Unambiguous condition suggested for the Written and RSVP conditions, the ambiguity effect represented a change from a low level of activity in this region in the Unambiguous conditions to a higher level of activity in the Ambiguous conditions. In the case of the Spoken condition, the ambiguity effect appeared to represent an increase from relative underactivity in the Unambiguous condition.

Discussion

The aim of Experiment 2 was to investigate the effects of presentation format-specific processing demands on higher-order neural language comprehension processes. Results from the post-scan Disambiguation test demonstrated that the participants were generally able to disambiguate the sentences, and had access to the relevant vocabulary to do so. The Disambiguation task confirmed the finding from Experiment 1 that sentences

that require ambiguity resolution are more difficult to process than sentences that do not contain ambiguities; even with prior exposure during scanning, participants showed lower accuracy rates in selecting the appropriate meaning for the ambiguous main noun than for the equivalent unambiguous noun in the control sentences. These results are in line with the striking findings that have led to the formulation of the "good enough" framework of language processing (see e.g. Christianson et al., 2001; Ferreira & Patson, 2007).

The fMRI results revealed ambiguity effects on haemodynamic responses in the Spoken, whole-sentence Written, and RSVP formats that were consistent with previous studies (e.g. Braze et al., 2011; Hoenig & Scheef, 2009; Rodd et al., 2012; Rodd et al., 2005; Vitello, 2014; Zhu et al., 2012). Importantly, these analyses demonstrated the critical finding that the magnitude of ambiguity-related processing costs in a key language region, left IFG, was modulated by the format in which sentences were presented. We observed that ambiguity effects in an anterior (BA 47) and a more posterior (BA 45) subregion of left IFG were significantly increased when the sentences were presented in the less naturalistic and more cognitively demanding word-by-word RSVP format.

General discussion

The present study investigated whether the processing demands of typically used presentation formats in psycho linguistic and neurolinguistic research affect higher-level semantic processes during language comprehension. In separate behavioural and neuroimaging experiments, processing of sentences requiring resolution of lexical-semantic ambiguity was compared to comprehension of unambiguous sentences in three commonly-used presentation formats (Spoken, whole-sentence Written, and RSVP). In both experiments, sentence stimuli for the ambiguous conditions were structured around late disambiguation to the subordinate meaning of an ambiguous key noun, so that successful comprehension was predicated primarily on reinterpretation of the sentence after encountering the sentence-terminal disambiguating word. Interactions between presentation format and the expected reinterpretation-related costs to behavioural measures during a meaning coherence judgement task and haemodynamic responses in ambiguity-responsive leftIFG regions during attentive comprehension were investigated.

In line with previous studies of ambiguity resolution, reinterpretation-related processing costs were consistently found in all three presentation formats (Spoken, Written and RSVP), both at the behavioural and the

neural level (Vitello & Rodd, 2015). Explicit ambiguity resolution during a behavioural task (Experiment 1) was associated with significantly lower accuracy and longer response times compared to matched unambiguous sentences in all three formats. Implicit comprehension of sentences containing semantic ambiguities (Experiment 2) was associated with significantly greater haemodynamic responses in an anterior leftIFG region (BA47) and a more posterior leftIFG region (BA45) in all presentation formats, with the exception of the Spoken format in BA47 (which was statistically non-significant, $p = .053$, Cohen's $d = 0.51$); post-scan behavioural testing confirmed the same pattern of ambiguity-related behavioural effects on sentence comprehension accuracy and speed. It is worth noting that although all of our sentence stimuli were designed to disambiguate an ambiguous word towards an unexpected meaning (thereby triggering reinterpretation processes), we did not tightly control the degree to which our ambiguous key words were biased towards a single, dominant meaning. Extensive piloting suggested that our stimulus set included a mix of more or less strongly biased ambiguous words; it is therefore possible that participants successfully resolved the ambiguity in sentences with less strongly biased ambiguous words during a first-pass through the sentence (therefore eliminating the need for reinterpretation), simply by virtue of our targeted "subordinate" meaning being more generally accessible (for such effects of word-meaning bias in eye-tracking studies see e.g. Duffy et al., 1988; Sereno & Rayner, 1993). Comparisons between the unambiguous condition and our ambiguous condition therefore likely resulted in a relatively conservative estimate of the true costs of reinterpretation. Future studies of reinterpretation-related processing costs may increase statistical power by selecting strongly biased ambiguous words only, e.g. by using large-scale word-meaning dominance norms that have now become available (e.g. Gilbert & Rodd, 2022, for norms of British English).

Apart from providing important replications of previously observed reinterpretation-related processing costs, our key research question was whether the format in which linguistic information is encountered can modulate such higher-level semantic processes. Importantly, in both experiments, the magnitude of reinterpretation-related processing costs varied significantly between sentence presentation formats. In Experiment 1, the magnitude of the ambiguity-related cost to processing speed (though not accuracy) during explicit sentence comprehension was greater for both the RSVP and the Spoken formats relative to the whole-sentence Written format. In Experiment 2, the magnitude of the

ambiguity-related increase in neural activity during implicit sentence comprehension was greater for the RSVP format than for either of the other two formats within BA47, and was greater for the RSVP format compared to the Spoken format within BA45.

The functional organisation of leftIFG remains a controversial topic, with opinion divided between domain-specific (e.g. distinctions between phonology, syntax and semantics, see Friederici, 2012; Grodzinsky, 2000) and process-specific (e.g. top-down control processes related to conflict resolution or more general executive functions, see Bookheimer, 2002; Curtis & D'Esposito, 2003; Kaan & Swaab, 2003; Nozari & Thompson-Schill, 2016; Postle, 2006; Thompson-Schill et al., 1997) models of left IFG function. While the two leftIFG regions demonstrated somewhat different patterns of interaction between presentation format and ambiguity, the absence of a significant interaction with region meant that a clear differentiation between the functional profiles of these two regions could not be made.

Although relative differences between formats were not identical in the behavioural and neuroimaging experiments, the results from both experiments indicate that it was the less naturalistic RSVP format which had the greatest impact on sentence comprehension in general, and semantic ambiguity resolution in particular. In Experiment 1, the RSVP format was associated with reduced accuracy rates for all sentence types (ambiguous and unambiguous) compared to the other presentation formats. This is consistent with previous work demonstrating that word-by-word presentation renders comprehension more difficult or effortful during explicit and implicit sentence comprehension (e.g. Masson, 1983; Schotter et al., 2014; Wlotko & Federmeier, 2015). The RSVP format was also associated with a significantly greater impact on the processing costs of ambiguity resolution compared to conventional whole-sentence Written presentation, both in terms of behavioural response times in Experiment 1 and neural activity within BA47 in Experiment 2.

Unlike whole-sentence Written presentation, the RSVP format precludes the use of visual input as an external working-memory buffer (Lee & Newman, 2009) and prevents overt re-reading as a means to aid reinterpretation of ambiguous sentences. Eye-tracking research has shown that readers tend to make eye movements towards earlier portions of a sentence when faced with processing difficulties (e.g. Frazier & Rayner, 1982; Paape & Vasissth, 2022). Evidence for increased backwards-directed eye movements has been found in particular for stimuli where readers are faced with unexpected information that disambiguates an earlier ambiguous sentence element, suggesting that physical

re-reading may play an important role in the successful recovery from misinterpretations during reading (e.g. Blott et al., 2021; Jacob & Felsler, 2016; Pickering & Traxler, 1998; Slattery et al., 2013). Resolving semantic ambiguity in the RSVP format may therefore place particular strain on working memory, as memory externalisation (i.e. relying on physical re-reading) is not possible. In addition, the RSVP format is associated with little, if any, prior practice and exposure – in contrast to whole-sentence reading, which is a highly practiced everyday skill from childhood for typical adults. It would therefore be surprising if the unfamiliar RSVP format did not place additional cognitive demands on the reader. We argue that, compared to whole-sentence reading, the RSVP format places greater demands on working memory (likely involving both storage and processing components), not only during sentence reading itself, but also during reinterpretation to resolve semantic ambiguity. The increased ambiguity-related behavioural and neural processing costs observed in the present study when visual sentences were presented word-by-word rather than as whole sentences may therefore reflect the enhanced impact of format-related working memory and executive control demands on sentences that require reinterpretation to achieve successful comprehension.

In addition to working memory demands, other format-related differences in cognitive aspects of sentence comprehension may have contributed to the increased ambiguity costs observed with RSVP presentation. The ambiguous sentences used in the present study incorporated a considerable structural and temporal distance between the ambiguous noun and the disambiguating sentence-terminal word, and may thus have encouraged “digging-in” effects (see e.g. Hagoort, 2003; Metzner et al., 2017). Such effects occur when comprehenders become increasingly committed to their initial analysis, and find it harder to re-analyse a sentence (Tabor & Hutchins, 2004). When words are presented sequentially, digging-in effects may become exaggerated compared to whole-sentence presentation during reading (Magliano et al., 1993). Mechanisms that are available during natural reading but lacking during RSVP reading, such as parafoveal preview and the ability to make regressive eye movements to previous parts of the sentence, could lead to an exaggeration of sentence wrap-up effects at the offset of Ambiguous sentences. In the present study, reinterpretation could only occur after the disambiguating sentence-terminal word; to capture this reinterpretation process behavioural response times (in Experiment 1) and neural activation (in Experiment 2) were recorded from sentence offset, meaning that measures of disambiguation

coincided with post-sentence wrap-up. The significant increase in ambiguity effects in the RSVP format compared to naturalistic whole-sentence reading might therefore reflect an interaction between format-related post-sentence wrap-up effects and the specific demands for reinterpretation posed by late-disambiguating sentences. In the case of the increased ambiguity-related neural activity observed in BA47 with RSVP, this explanation fits in particularly well with Hagoort’s (2005) account of left IFG function, which views anterior left IFG as fundamental to the process of binding together word meanings, and forming unified representations of sentence meaning.

Although the comparisons between RSVP and whole-sentence Written formats provide evidence supporting the idea that format-induced cognitive differences can influence the processing costs of ambiguity resolution, additional evidence from comparisons between the RSVP and Spoken formats suggests that task-related factors can also interact with format effects. RSVP mimics the transient nature of speech. Processing of RSVP sentences therefore might be expected to show a similar response pattern to the Spoken format if the cognitive demands of dynamic stimuli were the primary determinant of processing costs. This was indeed what we observed for the behavioural costs of sentence reinterpretation during the explicit comprehension task in Experiment 1, where ambiguity-related effects did not differ significantly between these two formats. However, the relative effects of these two formats on ambiguity-related neural activation in both left IFG ROIs during implicit sentence comprehension in Experiment 2 followed a different pattern; compared to RSVP, the Spoken format was associated with significantly smaller ambiguity-related increases in haemodynamic responses in BA 45 and BA47. It is possible that this finding reflects the use of continuous rather than sparse MRI sequences to enable modelling of haemodynamic responses specifically time-locked to the encounter of disambiguating information; the presence of scanner noise may have systematically affected comprehensibility of the Spoken condition, minimising the difference in haemodynamic responses to ambiguous and unambiguous sentences. However, each participant’s ability to hear and understand auditory sentence stimuli clearly over scanner noise was checked prior to data acquisition, and Recognition Memory Test accuracy scores for Spoken sentences were significantly above chance. The discrepancy between RSVP and Spoken format effects that we observed depending on the nature of the comprehension paradigm (explicit vs implicit) may also have a more theoretically-interesting, cognitive explanation. When comprehension difficulties

are encountered in the absence of task-related contexts or objectives to facilitate engagement with comprehension, comprehenders may tend to adopt “good-enough” sentence processing strategies (e.g. Christianson et al., 2001; Ferreira, 2003; Ferreira & Patson, 2007), whereby sentence processing remains at a shallow or superficial level, without representations of sentence meaning being fully specified or ambiguities fully resolved.

Greater reliance on good-enough processing during listening could explain the relatively reduced ambiguity effects found in the Spoken condition. It has previously been argued that such strategies of underspecification of sentence-level meaning representations may be more likely during listening than reading due to the associated time pressure for comprehension of transient speech stimuli (Sanford & Sturt, 2002). Although there was no explicit task during scanning to impose time pressures on comprehension, the pacing of stimulus delivery essentially constrained the amount of time participants could devote to comprehension of one sentence before the next was presented. Superficial, underspecified processing of ambiguous sentences would be expected to reduce engagement in sentence reinterpretation, limiting the amount of additional neural activation observed in response to ambiguous sentences. This was the response pattern we observed for the Spoken format in the BA47 ROI, where the difference in activation between Ambiguous and Unambiguous Spoken sentences was only of borderline significance. The same conditions of stimulus transience and time-limited processing should also apply to the RSVP sentences, so that underspecified processing of sentence meaning, and its consequences on neural activation, might also be expected for the RSVP Ambiguous sentences. However, the magnitude of the ambiguity-related increase in activation observed in Experiment 2 was significantly larger for the RSVP format compared with the Spoken format in both leftIFG ROIs, suggesting a much greater level of engagement with processing Ambiguous RSVP sentences. It seems plausible that factors such as the relative unfamiliarity or artificiality of RSVP sentence presentation, the increased working memory demands of serial word-by-word presentation in this format, or the relative salience of seeing the unexpected disambiguating word in isolation at the end of the RSVP word sequence may have been sufficient to drive more active engagement in comprehension of ambiguous sentences beyond “good-enough” processing, even in the absence of an explicit task, in a way that the more familiar and less attentionally engaging Spoken format did not. Taken together, the results from both experiments suggest that ambiguity-related processing costs may vary, not only due to format-

related differences in cognitive demands (e.g. working memory load, executive control demands) during sentence processing, but also depending on the extent that task characteristics (such as explicit vs implicit comprehension) interact with presentation format to encourage or hamper engagement with full specification of sentence meaning.

While the results of the present study support the existence of format-driven differences in ambiguity resolution, it remains possible that more prosaic design and analysis elements contributed to the differences in reinterpretation-related processing costs we observed between formats. One factor to consider is the difference in measurement timing necessitated by the characteristics of the different format conditions. Behavioural response times and haemodynamic responses were both measured from sentence offset, in order to time-lock measurements to the presentation of disambiguating information. In the Spoken and RSVP conditions, the nature of stimulus presentation meant that sentence offset could be precisely identified for all participants; this was not possible for the Written sentences, which necessitated the use of an estimate instead. For some participants and trials, these estimates may have placed Ambiguous sentence offset either too early (before sentence reinterpretation had begun) or too late (when sentence reinterpretation was already partially completed); either type of mis-timing would serve to weaken any observed ambiguity processing costs. In line with this explanation, response times in Experiment 1 tended to be shorter overall for the Written condition compared to the other two presentation formats, suggesting that sentence reading may have been completed earlier than estimated. Despite the finding of significant ambiguity-related increases in response times and neural activation for the Written conditions in both experiments, we cannot exclude the possibility that uncertainties of sentence offset timing could have contributed to the differences in magnitude of ambiguity-related processing costs we observed between the RSVP and Written conditions. To answer these questions more definitively in future, simultaneously collected eye-movement data both inside and outside the scanner could usefully be related to behavioural and neural measures. However, it is worth noting that issues of response timing in the present study do not impact upon the reliability of the comparisons between the RSVP and Spoken formats, since sentence offset could be identified precisely in both these formats.

Like the previous study by Lee and Newman (2009) investigating syntactic complexity effects, the present study provides evidence in support of the general

premise that presentation format can influence the behavioural and neural costs of processing difficult-to-interpret sentences. Furthermore, the present study expands this premise to show that format effects on processing costs are not confined to syntactic processing, but also extend to higher-level semantic processing. The Lee & Newman study and the present study demonstrate similar relationships between format and the behavioural costs of enhanced sentence processing demands, with both studies demonstrating greater detrimental effects of RSVP format on behavioural measures of explicit sentence comprehension. However, the relationships between neural processing costs and presentation format differ markedly between the two studies. The present study demonstrated larger ambiguity-related effects on neural activation within anterior leftIFG (BA47) for RSVP, in comparison to the whole-sentence Written format. In contrast, Lee and Newman (2009) found that whole-sentence written presentation was associated with a significant effect of syntactic complexity on posterior LIFG (BA44) activation, while the RSVP format was not. Once again, the explanation for this lack of consistency may rest with differences between the experimental tasks used in each study. In the Lee & Newman study, participants performed an explicit comprehension task during scanning, making true/false judgements on sentence probes presented after target sentences, and the presentation speed of the RSVP condition was relatively slow (about 400 ms per word); both these factors likely increased demands on working memory and executive control processes during RSVP sentence processing. Increased haemodynamic responses in the RSVP condition were observed in control sentences as well as the syntactically complex sentences, but error rates were higher for syntactically complex RSVP sentences, suggesting that syntactic complexity was not always successfully resolved despite this additional recruitment of neural resources. In the present study, there was no overt task during scanning and RSVP presentation speed was faster (250 ms per word), and, as discussed above, the pattern of activation in the RSVP condition suggests the possibility that under these circumstances, the cognitive exigencies of processing RSVP sentences may have served to selectively *increase* engagement with ambiguity resolution.

Regardless of the precise explanation for format-related differences in processing costs observed within the present study and in comparison with previous research, the existence of these differences is an important finding. It suggests that effects of presentation format on the costs of processing difficult-to-interpret sentences may vary considerably depending on the

nature of the cognitive demands placed on the comprehender by the format itself or the nature of the comprehension task. In particular, effects of linguistic manipulations observed in the RSVP format may not be reliably similar in magnitude to those observed in more naturalistic whole-sentence written or spoken formats. It remains unclear whether these format-related differences in higher-level semantic processing are driven by merely quantitative or substantial, qualitative differences in processing, such as the involvement of different cognitive processes, brain regions and neural pathways at different time scales, or different task strategies. More fine-grained investigations taking into account individual differences in functional neuroanatomy may be able to answer this question in future (Juch et al., 2005; Tahmasebi et al., 2012). For example, approaches using functional localisers may be able to find Region \times Format interactions within leftIFG subregions, providing more substantial evidence on whether the greater discrepancy between easy and complex processing found for the RSVP format is driven by quantitatively different engagement of the language network, or by additional engagement of other networks, e.g. the multiple demand network (Blank & Fedorenko, 2020; Diachek et al., 2020; Duffau, 2017; Quillen et al., 2021). In addition to increasing the sensitivity of fMRI measures to manipulations-of-interest compared to group-level analyses, such localiser approaches also have the benefit of avoiding problems with reverse inference (Nieto-Castañón & Fedorenko, 2012; Poldrack, 2006; Saxe et al., 2006).

The finding that format effects exist at all has important practical implications for language researchers, given that RSVP has been the dominant method to study neural responses during comprehension of connected text in fMRI, EEG and MEG studies because it allows precise time-locking of neural responses to the linguistic events being studied (see e.g. review on neuroscientific studies on reading by Caplan, 2004, and a more recent review of neuroimaging studies on reading syntactically complex sentences by Grodzinsky et al., 2021). In future, it will therefore be important for researchers to consider the RSVP format's potentially exaggerated sensitivity to manipulations of semantic processing, and explicitly take such characteristics into account when designing studies. An avenue for investigating higher-level language processing in the written modality while avoiding the potential pitfalls of RSVP discussed here may be to combine more naturalistic presentation of sentence stimuli (i.e. as whole-sentence written input) with techniques which allow for time-locked measurement of haemodynamic responses to specific linguistic events during self-paced reading,

such as a combination of fMRI with eye-tracking (see e.g. Bonhage et al., 2015; Henderson et al., 2016; Schuster et al., 2015, 2020).

Conclusions

How linguistic stimuli are presented matters: this study has demonstrated that the use of RSVP, whole-sentence Written and Spoken sentence formats produces differential effects on the processing costs of sentence reinterpretation during semantic ambiguity resolution. However, the present study cannot conclusively determine whether the modality-related differences we observed are driven by theoretically important cognitive differences in sentence processing, e.g. because different processing formats load differently on factors such as working memory, or encourage differences in the extent to which participants engage in “good enough” processing or “digging in”. Further studies which systematically vary the properties of the stimuli or the task are needed to test specific predictions about how the cognitive/neural mechanisms supporting sentence comprehension are influenced by presentation formats.

Although our findings support the idea that format-related – and task-related cognitive demands contribute to these differential effects on higher-level language comprehension, the pattern of results from this and previous studies suggests that predicting exactly how a given presentation format will affect higher-level language processing may not be straightforward. The present study highlights the need for better understanding of the processing demands of different presentation formats for linguistic material, and greater awareness of the potential impact of presentation format for both behavioural and neuroimaging research into higher-level sentence comprehension processes such as sentence reinterpretation.

Acknowledgements

This work was part of the first author’s doctoral dissertation at University College London, which was supported by an ESRC-MRC Multidisciplinary Studentship to LMB. An earlier version of this work was included in LMB’s PhD thesis submitted in 2019. The authors would like to thank Bonnie Kim for assistance with the experiment set up and data collection for Experiment 1.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Economic and Social Research Council; Medical Research Council (MRC).

Data availability statement

The data that support the findings of Experiment 1 and anonymised summary data that support the findings of Experiment 2 are openly available at the Open Science Framework at <https://osf.io/m87vg/>. The raw data that support the findings of Experiment 2 of this study are available on request from the senior author, JEW. These data are not publicly available due to their containing information that could compromise the privacy of research participants.

ORCID

Lena M. Blott  <http://orcid.org/0000-0001-5268-0763>

References

- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52, 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, 26(3), 839–51. <https://doi.org/10.1016/j.neuroimage.2005.02.018>
- 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>
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (release 2)*. Distributed by the Linguistic Data Consortium, University of Pennsylvania.
- 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., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious Mixed Models*. ArXiv:1506.04967 [Stat]. <http://arxiv.org/abs/1506.04967>
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12), 2767–2796. <https://doi.org/10.1093/cercor/bhp055>
- Blank, I. A., & Fedorenko, E. (2020). No evidence for differences among language regions in their temporal receptive windows. *NeuroImage*, 219, 116925.
- Blessner, B. (1972). Speech Perception Under Conditions of Spectral Transformation: I. Phonetic Characteristics. *Journal of Speech and Hearing Research*, 15(1), 5–41. <https://doi.org/10.1044/jshr.1501.05>
- Blott, L. M., Rodd, J. M., Ferreira, F., & Warren, J. E. (2021). Recovery from misinterpretations during online sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(6), 968–997. <https://doi.org/10.1037/xlm0000936>

- Bonhage, C. E., Mueller, J. L., Friederici, A. D., & Fiebach, C. J. (2015). Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex*, 68, 33–47. <https://doi.org/10.1016/j.cortex.2015.04.011>
- Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25(1), 151–188. <https://doi.org/10.1146/annurev.neuro.25.112701.142946>
- Braze, D., Mencl, W. E., Tabor, W., Pugh, K. R., Todd Constable, R., Fulbright, R. K., ... Shankweiler, D. P. (2011). Unification of sentence processing via ear and eye: An fMRI study. *Cortex*, 47(4), 416–431. <https://doi.org/10.1016/j.cortex.2009.11.005>
- Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002, June). Region of interest analysis using an SPM toolbox. In *8th International Conference on Functional Mapping of the Human Brain*, 16(2), 497.
- Caplan, D. (2004). Functional neuroimaging studies of written sentence comprehension. *Scientific Studies of Reading*, 8(3), 225–240. https://doi.org/10.1207/s1532799xssr0803_3
- 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>
- Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. J. S. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain*, 126(5), 1193–1201. <https://doi.org/10.1093/brain/awg104>
- Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, 7(9), 415–423. [https://doi.org/10.1016/S1364-6613\(03\)00197-9](https://doi.org/10.1016/S1364-6613(03)00197-9)
- Damer, E., & Bradley, P. (2014). Prolific Academic: Prolific Academic Ltd. (prolific.ac). Available at www.prolific.ac
- Diachek, E., Blank, I., Siegelman, M., Affourtit, J., & Fedorenko, E. (2020). The domain-general multiple demand (MD) network does not support core aspects of language comprehension: a large-scale fMRI investigation. *Journal of Neuroscience*, 40(23), 4536–4550.
- Duffau, H. (2017). A two-level model of interindividual anatomo-functional variability of the brain and its implications for neurosurgery. *Cortex*, 86, 303–313.
- 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](https://doi.org/10.1016/0749-596X(88)90066-6)
- Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, 108(39), 16428–16433. <https://doi.org/10.1073/pnas.1112937108>
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2), 164–203. [https://doi.org/10.1016/S0010-0285\(03\)00005-7](https://doi.org/10.1016/S0010-0285(03)00005-7)
- 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>
- 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](https://doi.org/10.1016/0010-0285(82)90008-1)
- Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16(5), 262–268. <https://doi.org/10.1016/j.tics.2012.04.001>
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 189–210. <https://doi.org/10.1002/hbm.460020402>
- Gilbert, R. A., & Rodd, J. M. (2022). Dominance norms and data for spoken ambiguous words in British English. *Journal of Cognition*, 5(1).
- Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, 23(1), 1–21. <https://doi.org/10.1017/S0140525X00002399>
- Grodzinsky, Y., Pieperhoff, P., & Thompson, C. (2021). Stable brain loci for the processing of complex syntax: A review of the current neuroimaging evidence. *Cortex*, 142, 252–271. <https://doi.org/10.1016/j.cortex.2021.06.003>
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15(6), 883–899. <https://doi.org/10.1162/089892903322370807>
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9(9), 416–423. <https://doi.org/10.1016/j.tics.2005.07.004>
- Henderson, J. M., Choi, W., Lowder, M. W., & Ferreira, F. (2016). Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in Reading. *Neuroimage*, 132, 293–300. <https://doi.org/10.1016/j.neuroimage.2016.02.050>
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1–2), 67–99. <https://doi.org/10.1016/j.cognition.2003.10.011>
- Hoenig, K., & Scheef, L. (2009). Neural correlates of semantic ambiguity processing during context verification. *NeuroImage*, 45(3), 1009–1019. <https://doi.org/10.1016/j.neuroimage.2008.12.044>
- Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology*, 69(5), 907–925. <https://doi.org/10.1080/17470218.2014.984231>
- JASP Team. (2019). JASP (Version 0.10.2)[Computer software].
- Juch, H., Zimine, I., Seghier, M. L., Lazeyras, F., & Fasel, J. H. (2005). Anatomical variability of the lateral frontal lobe surface: implication for intersubject variability in language neuroimaging. *Neuroimage*, 24(2), 504–514.
- Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15(1), 98–110. <https://doi.org/10.1162/089892903321107855>
- 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>
- Lee, D., & Newman, S. D. (2009). The effect of presentation paradigm on syntactic processing: An event-related fMRI study. *Human Brain Mapping*, NA–NA. <https://doi.org/10.1002/hbm.20845>

- Magliano, J. P., Baggett, W. B., Johnson, B. K., & Graesser, A. C. (1993). The time course of generating causal antecedent and causal consequence inferences. *Discourse Processes*, 16(1–2), 35–53. <https://doi.org/10.1080/01638539309544828>
- Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38(3), 487–497. [https://doi.org/10.1016/S0896-6273\(03\)00197-1](https://doi.org/10.1016/S0896-6273(03)00197-1)
- Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research*, 1146, 115–127. <https://doi.org/10.1016/j.brainres.2007.02.076>
- Masson, M. E. J. (1983). Conceptual processing of text during skimming and rapid sequential Reading. *Memory & Cognition*, 11(3), 262–274. <https://doi.org/10.3758/BF03196973>
- Melchert, C. M. C. (2004). In: Woodward, R. D. (Ed.). *The Cambridge Encyclopaedia of the World's Ancient Languages*. Cambridge: Cambridge University Press.
- Mesulam, M. (1998). From sensation to cognition. *Brain*, 121(6), 1013–1052. <https://doi.org/10.1093/brain/121.6.1013>
- Metzner, P., von der Malsburg, T., Vasisht, S., & Rösler, F. (2017). The importance of Reading naturally: Evidence from combined recordings of eye movements and electric brain potentials. *Cognitive Science*, 41, 1232–1263. <https://doi.org/10.1111/cogs.12384>
- Nieto-Castañón, A., & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage*, 63(3), 1646–1669. <https://doi.org/10.1016/j.neuroimage.2012.06.065>
- Nozari, N., & Thompson-Schill, S. L. (2016). Left ventrolateral prefrontal cortex in processing of words and sentences. In *Neurobiology of language* (pp. 569–584). <https://doi.org/10.1016/B978-0-12-407794-2.00046-8>.
- Paape, D., & Vasisht, S. (2022). Is reanalysis selective when regressions are consciously controlled? *Glossa Psycholinguistics*, 1(1), <https://doi.org/10.5070/G601139>
- Pacht, J. M., & Rayner, K. (1993). The processing of homophonic homographs during Reading: Evidence from eye movement studies. *Journal of Psycholinguistic Research*, 22(2), 251–271. <https://doi.org/10.1007/BF01067833>
- 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. <https://doi.org/10.1037/0278-7393.24.4.940>
- Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10(2), 59–63. <https://doi.org/10.1016/j.tics.2005.12.004>
- Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, 139(1), 23–38. <https://doi.org/10.1016/j.neuroscience.2005.06.005>
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, 62(2), 816–847. <https://doi.org/10.1016/j.neuroimage.2012.04.062>
- Quillen, I. A., Yen, M., & Wilson, S. M. (2021). Distinct neural correlates of linguistic and non-linguistic demand. *Neurobiology of Language*, 2(2), 202–225.
- Rodd, J. (2018). *Lexical ambiguity*. Oxford handbook of psycholinguistics. 120–144.
- Rodd, J. M. (2020). Settling into semantic space: An ambiguity-focused account of word-meaning access. *Perspectives on Psychological Science*, 15, 411–427. <https://doi.org/10.1177/1745691619885860>
- Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, 15(8), 1261–1269. <https://doi.org/10.1093/cercor/bhi009>
- 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(8), 1761–1773. <https://doi.org/10.1093/cercor/bhr252>
- Rodd, J. M., Longe, O. A., Randall, B., & Tyler, L. K. (2010). The functional organisation of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia*, 48(5), 1324–1335. <https://doi.org/10.1016/j.neuropsychologia.2009.12.035>
- sRodd, J. M., Vitello, S., Devlin, J. T., & Warren, J. E. (2013, November 6–8). The roles of left and right inferior frontal cortex in the comprehension of ambiguous sentences [Conference abstract]. *Society for the Neurobiology of Language 2013 Annual Meeting*, San Diego, CA, United States.
- RStudio Team (2015). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc. <http://www.rstudio.com/>
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6(9), 382–386. [https://doi.org/10.1016/S1364-6613\(02\)01958-7](https://doi.org/10.1016/S1364-6613(02)01958-7)
- Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *Neuroimage*, 30(4), 1088–1096. <https://doi.org/10.1016/j.neuroimage.2005.12.062>
- Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what You read (only once). *Psychological Science*, 25(6), 1218–1226. <https://doi.org/10.1177/0956797614531148>
- Schuster, S., Hawelka, S., Himmelstoss, N. A., Richlan, F., & Hutzler, F. (2020). The neural correlates of word position and lexical predictability during sentence Reading: Evidence from fixation-related fMRI. *Language, Cognition and Neuroscience*, 35(5), 613–624. <https://doi.org/10.1080/23273798.2019.1575970>
- Schuster, S., Hawelka, S., Richlan, F., Ludersdorfer, P., & Hutzler, F. (2015). Eyes on words: A fixation-related fMRI study of the left occipito-temporal cortex during self-paced silent Reading of words and pseudowords. *Scientific Reports*, 5(1), 1–11. <https://doi.org/10.1038/srep12686>
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123(12), 2400–2406.
- Sereno, S. C., & Rayner, K. (1993, November). *Lexical ambiguity resolution: Evidence from an eye movement priming paradigm*. Washington, DC: Paper presented at the 34rd Annual Meeting of the Psychonomic Society.
- 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](https://doi.org/10.1016/S0022-5371(81)90356-X)
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Linger misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, 69(2), 104–120. <https://doi.org/10.1016/j.jml.2013.04.001>

- Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. S. (2006). Converging language streams in the human temporal lobe. *Journal of Neuroscience*, 26(28), 7328–7336. <https://doi.org/10.1523/JNEUROSCI.0559-06.2006>
- 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](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>
- Tahmasebi, A. M., Davis, M. H., Wild, C. J., Rodd, J. M., Hakyemez, H., Abolmaesumi, P., & Johnsrude, I. S. (2012). Is the link between anatomical structure and function equally strong at all cognitive levels of processing?. *Cerebral cortex*, 22(7), 1593–1603.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, 94(26), 14792–14797. <https://doi.org/10.1073/pnas.94.26.14792>
- 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>
- Tyler, L. K., & Marslen-Wilson, W. (2008). Fronto-temporal brain systems supporting spoken language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1037–1054. <https://doi.org/10.1098/rstb.2007.2158>
- Vitello, S. (2014). *Cognitive and Neural Mechanisms Underlying Semantic Ambiguity Resolution*. (Doctoral dissertation, University of London).
- Vitello, S., & Rodd, J. M. (2015). Resolving semantic ambiguities in sentences: Cognitive processes and brain mechanisms. *Language and Linguistics Compass*, 9(10), 391–405. <https://doi.org/10.1111/lnc3.12160>
- 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>
- Warren, J. E. (2013). *The functional organisation of neural systems supporting human vocal communication*. (Doctoral dissertation, University of London).
- Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of just and carpenter (1992). *Psychological Review*, 103(4), 761–772. <http://dx.doi.org/10.1037/0033-295X.103.4.761>
- Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word Reading. *Cortex*, 68, 20–32. <https://doi.org/10.1016/j.cortex.2015.03.014>
- Wright, P., Randall, B., Marslen-Wilson, W. D., & Tyler, L. K. (2011). Dissociating linguistic and task-related activity in the left inferior frontal gyrus. *Journal of Cognitive Neuroscience*, 23(2), 404–413. <https://doi.org/10.1162/jocn.2010.21450>
- Zempleni, M.-Z., Renken, R., Hoeks, J. C. J., Hoogduin, J. M., & Stowe, L. A. (2007). Semantic ambiguity processing in sentence context: Evidence from event-related fMRI. *NeuroImage*, 34(3), 1270–1279. <https://doi.org/10.1016/j.neuroimage.2006.09.048>
- Zhu, Z., Hagoort, P., Zhang, J. X., Feng, G., Chen, H.-C., Bastiaansen, M., & Wang, S. (2012). The anterior left inferior frontal gyrus contributes to semantic unification. *NeuroImage*, 60(4), 2230–2237. <https://doi.org/10.1016/j.neuroimage.2012.02.036>