

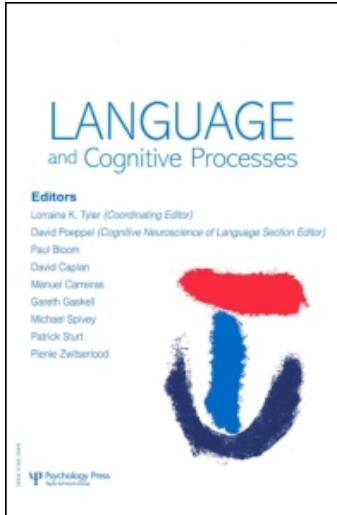
This article was downloaded by: [Tufts University]

On: 5 April 2011

Access details: Access Details: [subscription number 930565352]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Language and Cognitive Processes

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713683153>

The contributions of lexico-semantic and discourse information to the resolution of ambiguous categorical anaphors

Tali Ditman^a; Phillip J. Holcomb^a; Gina R. Kuperberg^{ab}

^a Tufts University, Medford, MA, USA ^b Massachusetts General Hospital, Charlestown, MA, USA

First published on: 01 April 2007

To cite this Article Ditman, Tali, Holcomb, Phillip J. and Kuperberg, Gina R. (2007) 'The contributions of lexico-semantic and discourse information to the resolution of ambiguous categorical anaphors', *Language and Cognitive Processes*, 22: 6, 793 – 827, First published on: 01 April 2007 (iFirst)

To link to this Article: DOI: 10.1080/01690960601057126

URL: <http://dx.doi.org/10.1080/01690960601057126>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

The contributions of lexico-semantic and discourse information to the resolution of ambiguous categorical anaphors

Tali Ditman and Phillip J. Holcomb

Tufts University, Medford, MA, USA

Gina R. Kuperberg

*Tufts University, Medford, MA, and Massachusetts General Hospital,
Charlestown, MA, USA*

The present studies employed event-related potentials (ERPs) to examine the time course for the integration of lexico-semantic and discourse information during the resolution of categorical anaphors. Scenarios were constructed to include three potential antecedents. Anaphors were semantically ambiguous in that two of the potential antecedents were exemplars of the anaphor. Final sentences resolved the anaphor with the correct (associatively related/contextually appropriate), incorrect (associatively related/contextually inappropriate), or control antecedent (associatively unrelated/contextually inappropriate). We examined the amplitude of the N400 component, which is thought to reflect the ease of semantic integration, at several points following the anaphor. The smallest N400 was evoked when the text referred back to a correct antecedent following an anaphor; an intermediate N400 was evoked by incorrect antecedents and the largest N400 was evoked by reinstating the control antecedent following an anaphor. Results demonstrated that, following an ambiguous anaphor, readers are able to use both lexico-semantic and discourse-level information to semantically integrate an antecedent into its larger discourse context.

Correspondence should be addressed to Tali Ditman, M.A., Tufts University, 490 Boston Avenue, Medford, MA 02155; e-mail: Tali.Ditman@tufts.edu

This research was supported by NICHD grants HD25889 and HD043251 to Phillip J. Holcomb and by NIMH RO1 MH 071635 to Gina R. Kuperberg. In addition, Dr Kuperberg was supported by National Alliance for Research on Schizophrenia and Depression (through the Sidney Baer Trust), a Claflin Distinguished Scholars Award from Massachusetts General Hospital, and the Institute for Mental Illness and Neuroscience Discovery (MIND).

© 2007 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business

<http://www.psypress.com/lcp>

DOI: 10.1080/01690960601057126

Integrating incoming information into a larger discourse is an essential aspect of language comprehension. Discourse is differentiated from isolated sentences by two properties: coherence and cohesion (e.g., Halliday & Hasan, 1976; Sanford & Garrod, 1994). Coherence refers to the establishment of logical and psychological consistency between events (e.g., through inferences). Cohesion is established through linguistic devices that link information in sentences. One way to establish cohesion in a discourse is through the use of anaphors. An anaphor is a word or group of words that refers to a previously mentioned word or concept, the antecedent (e.g., Garrod & Sanford, 1994). The process of determining the best matching antecedent for an anaphor is termed anaphor resolution.

Although anaphor resolution is a common process during everyday language comprehension, it is not without demands. An anaphor often has several possible antecedents (e.g., the word 'he' could refer to one of several previously mentioned male characters in a story). The challenge for the comprehension system is to resolve this ambiguity, specifically linking an anaphor with its correct antecedent (e.g., Haviland & Clark, 1974). To do this, we rely on the surrounding sentence and/or discourse context. Despite the many behavioural studies devoted to the study of anaphor resolution (e.g., Corbett & Chang, 1983; Gernsbacher, 1989; Greene, McKoon, & Ratcliff, 1992; Lucas, Tanenhaus, & Carlson, 1990; O'Brien, Duffy, & Myers, 1986; Wiley, Mason, & Myers, 2001; see also Garnham, 2001, for a review), the precise neurocognitive processes involved remain unclear.

Gernsbacher (1989) postulated that anaphor resolution involves both the facilitation of a correct antecedent and the inhibition of competitors. For example, she reported that, in sentences with two female characters, when the anaphor 'she' referred to one of the characters, participants were slower to respond to competitor probes and faster to respond to correct antecedents following the anaphor compared with activation levels (i.e., as assessed by RTs) of these words preceding the anaphor. However, subsequent behavioural studies yielded equivocal results. Some studies have produced evidence for only facilitation (e.g., Nicol, 1988, unpublished doctoral dissertation cited in MacDonald & MacWhinney, 1990), others for only inhibition (e.g., MacDonald & MacWhinney, 1990; Wiley et al., 2001), and others still for both facilitation and inhibition (e.g., Gernsbacher, 1989).

In attempting to determine the source of these contradictory findings, Lucas et al. (1990) provided evidence suggesting that the dependent measure employed (i.e., naming times, lexical decision times) could account for discrepancies between experiments. When naming times (arguably sensitive to lexical but not post-lexical integrative processes) were used as the dependent measure, Lucas and colleagues found evidence only for facilitation of correct and incorrect antecedents that were both semantically related to the preceding anaphor. On the other hand, using a lexical decision task

(sensitive to post-lexical as well as lexical processes), inhibition for an incorrect antecedent was demonstrated. The authors concluded that the facilitation indexed through naming times was a byproduct of local, lexico-semantic processes whereas the inhibition indexed through lexical decision times was necessary for global, integrative processes associated with discourse comprehension and successful anaphor resolution.

The results from the above studies illustrate the importance of using an online measure that is not overly influenced by subjects' decisions. The present studies employed event-related potentials (ERPs) that allow for an online assessment of neural activity during language comprehension with millisecond (ms) temporal resolution, without requiring a behavioural response at the point of anaphor resolution. In addition, ERPs allow for a qualitative analysis of the data as well as a quantitative analysis. They give insights not only into *when* a process of interest occurs but also into the neurocognitive nature of this process. To examine the integration of semantic information into discourse context, we examined the amplitude of the N400 – an ERP component that is maximal over centroparietal sites, that peaks at approximately 400 ms after word onset, and that is thought to reflect the ease of semantically integrating a word into its preceding context (e.g., Holcomb, 1993; reviewed by Kutas & Federmeier, 2000).

Numerous studies have linked the N400 to lexico-semantic influences. Using typical semantic priming paradigms in which target words follow primes that are semantically related (e.g., doctor-nurse) or unrelated (e.g., dog-nurse), a larger N400 is observed to the semantically unrelated targets (e.g., Bentin, McCarthy, & Wood, 1985; Rugg, 1984).

The N400 is also known to be modulated by sentential factors. At the sentence level, classic studies by Kutas & Hillyard (1980, 1984) demonstrated that content words that are semantically incongruous with a sentence context elicit a larger N400 than words that are semantically congruous with the preceding context (e.g., She took her coffee with milk and socks/sugar). In addition, the amplitude of the N400 is sensitive to sentence constraint (as assessed by cloze probability), such that the amplitude of the N400 is smaller for words that are expected in a given context and larger for words that are unexpected, despite both expected and unexpected words being contextually appropriate.

More recent studies have provided evidence that the N400 is also sensitive to discourse-level factors (e.g., Anderson & Holcomb, 2005; Federmeier & Kutas, 1999; St George, Mannes, & Hoffman, 1994; Swaab, Camblin, & Gordon, 2004; Van Berkum, Hagoort, & Brown, 1999b; for a review see Van Berkum, 2004). St George et al. (1994) manipulated the coherence of a passage by providing a context-framing title (coherent condition) or by omitting this title (incoherent condition). Content words elicited a larger N400 in the incoherent condition, reflecting the sensitivity of the N400 to the

global context. In addition, in both visual and auditory modalities, Van Berkum and colleagues (in visual modality: Van Berkum et al., 1999b; in auditory modality: Van Berkum, Zwitserlood, Hagoort, & Brown, 2003) found that a critical word in a third sentence evoked a larger N400 when this sentence was incongruous than when it was congruous with information presented in preceding sentences. This modulation of the N400 could not be explained by sentence-level differences. In addition to its immediate sensitivity to lexico-semantic and sentential/discourse factors, the N400 amplitude is influenced by the structure of semantic memory (Federmeier & Kutas, 1999).

Interestingly, the waveform, scalp-distribution, and timing of the N400s evoked in word-, sentence-level and discourse paradigms are similar, suggesting that common neurocognitive processes may underlie both lexico-semantic and sentential priming (Kutas, 1993; Van Petten, 1993). It seems that the N400 is sensitive to a process that integrates semantic information across several levels: relationships between individual words, stored knowledge within semantic memory, and information from sentence- and discourse-level contexts.

In the current study we examined the role of semantic integration during the resolution of ambiguous anaphors. Most ERP studies examining cohesion within discourse have focused on the anaphor itself. Several studies have established that comprehenders may attempt to locate a correct antecedent for an anaphor as early as 280 ms after its presentation (Van Berkum, Brown, & Hagoort, 1999a). This is manifest by another ERP component – the left anterior negativity (LAN) that is thought to be sensitive to structure building syntactic processes (Friederici, 1995) as well as working memory demands during discourse comprehension (King & Kutas, 1995). Both the processes of identifying a correct antecedent for an anaphor as well as linking new information supplied in the anaphor to the representation of the antecedent (e.g., the anaphor ‘the robin’ adds new information about the antecedent ‘a bird’) may tax working memory (Almor, 1999).

Although these studies demonstrate that readers and listeners become aware of coreferentiality in a discourse soon after the presentation of a coreferent, comprehenders may not actually resolve the anaphor at this point. The current studies focused on the process of anaphor resolution. We measured ERPs as participants read five-sentence discourse scenarios. Each began with a three-sentence context that introduced one of three potential antecedents. The fourth sentence ended with a semantically ambiguous anaphor. Final sentences resolved the anaphor with a reinstatement of the antecedent that was either correct (associatively related/contextually appropriate), incorrect (associatively related/contextually inappropriate), or control (associatively unrelated/contextually inappropriate). We examined the

time course of semantic integration at the point of this reinstatement, focusing on how discourse-level and lexico-semantic information interact to correctly resolve the semantically ambiguous anaphor. Three experiments were conducted. In the first experiment (Experiments 1a and 1b), we validated the materials to be used in the main ERP experiment (Experiment 2). Experiment 2 was conducted in order to examine the influences of lexico-semantic and discourse factors in the resolution of ambiguous categorical NP anaphors. Finally, Experiment 3 was conducted in order to test alternative interpretations of Experiment 2 results.

STIMULUS CONSTRUCTION

Two hundred and fifty-two categories, each with two exemplars that were associated with strong contexts, were created. To avoid effects of item typicality demonstrated in previous studies (Almor, 1999; Duffy & Rayner, 1990; Garrod & Sanford, 1977; Levine, Guzman, & Klin, 2000; Van Gompel & Majid, 2004), items were rotated through conditions so that on different lists they served as correct, incorrect, and control antecedents. In order to achieve this, categories were paired together so that exemplars 1 and 2 of category 1 (e.g., seat: stool, couch) were paired with exemplar 1 (e.g., clock) or 2 (e.g., watch) of category 2 (e.g., timepiece). This resulted in the creation of four lists.

Five-sentence scenarios were constructed from the category/exemplar sets, and then randomly assigned to one of three reinstatement conditions: correct, incorrect, and control. Each of the first three sentences in a scenario presented one potential antecedent, which was always the subject of the sentence. These sentences were constructed so that two of the potential antecedents were exemplars of the anaphor. We will refer to the antecedents within these first three sentences as the correct antecedent, incorrect antecedent, and control antecedent. A correct antecedent was both locally semantically appropriate (i.e., associatively related to the anaphor) and globally, contextually appropriate (i.e., correct antecedent); an incorrect antecedent was locally appropriate (associatively related to the anaphor) but contextually inappropriate, and a control antecedent was neither appropriate at a local nor discourse level. The fourth sentence always ended with the anaphor and provided a sufficiently strong context to bias only one potential antecedent as the correct interpretation. The fifth sentence, referred to as the resolution sentence, began with a reinstatement of one of the three antecedents introduced in the first three sentences (with the first word always being 'The'). We will refer to the reinstated antecedents within the fifth sentence as the correct reinstatement, incorrect reinstatement, and control reinstatement. Four lists were developed, each with 126 scenarios, 42

from each of the three reinstatement conditions: correct, incorrect, and control. Each participant viewed one list, and therefore only saw one version of each scenario. The order of correct, incorrect, and control antecedent presentation was counterbalanced across scenarios such that each appeared in the first, second, and third sentences an equal number of times within each list.

Below is an example scenario from each of the four lists, demonstrating how items were rotated through each condition (although the examples below are shown with the correct antecedent presented first followed by the incorrect and control antecedents, the order in which these potential antecedents were presented was also rotated between-scenario). Antecedents appear in bold, anaphors appear in capital letters, and reinstatements are italicised:

1. A **stool** is found in a bar. A **couch** is found in a dining room. A **clock** is found on a wall. At the bar, Henry sat on the SEAT. The *stool_{correct}/couch_{incorrect}/clock_{control}* was recently purchased.
2. A **couch** is found in a dining room. A **stool** is found in a bar. A **clock** is found on a wall. In the dining room, Henry sat on the SEAT. The *couch_{correct}/stool_{incorrect}/clock_{control}* was recently purchased.
3. A **clock** is found on a wall. A **watch** is found on a wrist. A **stool** is found in a bar. Henry glanced at the wall to see the TIMEPIECE. The *clock_{correct}/watch_{incorrect}/stool_{control}* was recently purchased.
4. A **watch** is found on a wrist. A **clock** is found on a wall. A **couch** is found in a dining room. Henry glanced at his wrist to see the TIMEPIECE. The *watch_{correct}/clock_{incorrect}/couch_{control}* was recently purchased.

In order to ensure that observed differences between the conditions were not due purely to lexical differences, we matched all critical words on several variables. The two groups of categorical anaphors (i.e., two groups of 126) did not differ from each other in word frequency ($p > .10$; Kucera & Francis, 1967) and word length ($p > .10$). The four groups of exemplars (i.e., two per category) were also matched on word frequency ($p > .10$; Kucera & Francis, 1967) and word length ($p > .10$). Sentence final words for all three Reinstatement conditions were matched on word frequency ($p > .10$; Kucera & Francis, 1967) and word length ($p > .10$). On average, final sentences were 4.3 words in length ($SD = 0.76$) and the number of words did not differ between the conditions ($p > .10$ for all pairwise comparisons).

In order to objectively confirm that coherence was lower in the incorrect and control reinstatement conditions compared with the correct condition, each of the four lists was randomly assigned to one of four naive judges who were asked to read each scenario and rate how well the fifth sentence fits in

with the preceding information, on a scale from 1 to 7 with 1 being 'poorly coherent' and 7 being 'very coherent'. As expected, on average, scenarios with correct reinstatements were judged to be more coherent ($M = 5.88$, $SD = .84$) than scenarios with incorrect ($M = 2.27$, $SD = .54$) and control ($M = 2.25$, $SD = 1.03$) reinstatements, which did not differ from one another. These results were confirmed with paired *t*-tests (correct/incorrect: $p = .01$; correct/control: $p = .01$; incorrect/control: $p = .96$).

EXPERIMENT 1

We will first describe two preliminary studies (Experiments 1a and 1b) used to help validate the materials to be used in the main ERP experiment (Experiment 2). Experiment 1a was conducted to ensure that two antecedents (e.g., stool, couch) would be considered exemplars of the categorical anaphor (e.g., seat), and Experiment 1b was performed to ensure that, without the discourse context, there were no pre-existing differences between the final sentences of the three Reinstatement conditions.

Experiment 1a

For each scenario, it was important that two of the antecedents (e.g., stool, couch) would be considered exemplars of the categorical anaphor (e.g., seat) and, equally important, that one of the antecedents (e.g., clock) would not be considered an exemplar. To determine this, 19 Tufts University undergraduate students (all giving written informed consent) participated in a ratings study in which they read the three antecedent sentences (i.e., the first three sentences of each scenario) and identified which items, if any, were 'types' of the categorical anaphor. Each participant was presented with a seven-page packet of 168 scenarios: 126 experimental stimuli and 42 randomly interleaved filler items that were constructed with similar characteristics as the first three sentences of the experimental stimuli with the exception that either none of the items were exemplars of the category (14 fillers), one item was an exemplar of the category (14 fillers), or all three items were exemplars of the category (14 fillers). Each scenario contained three sentences. The three potential antecedents along with the categorical anaphor in bold were presented immediately below each set of three sentences. Participants indicated which, if any, of the three items were 'types' of the word in bold.

To examine ratings, for each participant, hit (HR) and false alarm (FAR) rates were computed. The HR was calculated by scoring each item that a participant correctly identified as a category exemplar as a '1'. Then, all correct responses were tallied and divided by the total number of items (252). FARs were calculated in a similar fashion by scoring all incorrect items that

participants had identified as category exemplars (e.g., indicating that clock was a type of seat) and dividing by the total number of distractor items (126). Participants consistently identified category exemplars ($M_{HR} = .95$, $SD_{HR} = .03$) and correctly rejected control antecedents (i.e., clock) ($M_{FAR} = .03$, $SD_{FAR} = .02$). To examine ratings by item type, we subtracted the HR from the FAR for each condition (i.e., correct, incorrect, and control). There were no differences between the three conditions ($M_{correct} = .92$, $SD_{correct} = .04$; $M_{incorrect} = .91$, $SD_{incorrect} = .04$; $M_{control} = .93$, $SD_{control} = .05$; $p > .10$ for all pairwise comparisons).

EXPERIMENT 1B

Experiment 1b was conducted to ensure that sentence-level differences did not exist between the three Reinstatement conditions without the preceding discourse. This was particularly important because our counterbalancing procedure resulted in different final sentences for each condition. The existence of sentence-level differences would make it difficult to interpret discourse influences on anaphor resolution. On the other hand, if differences are not found without the preceding context, then we can be confident that any differences observed with the context are not merely due to sentence-level effects. To this end, ERPs were recorded while participants read 'final' sentences (i.e., sentence 5) without the preceding discourse context (i.e., sentences 1–4).

Method

Participants

Sixteen native-English speakers with a mean age of 19.81 years (six male and ten female) participated in Experiment 1. Written informed consent was obtained from all participants in accordance with the guidelines of the Tufts Human Subjects Research Committee. All participants were right-handed, had no history of traumatic head injury, and had normal or corrected-to-normal vision.

Procedure

Participants sat approximately 76 cm away from a computer and read 126 'final' sentences. We use the same nomenclature (e.g., reinstatement, antecedent, resolution, final sentence, in quotation marks) for consistency even though, in this experiment, only one sentence was presented, without antecedents or anaphors. Participants were instructed to silently read all sentences for comprehension. Each trial began with the word 'READY' in the middle of the screen to give participants time to blink. To begin each

trial, participants pressed a button on a game-pad with their left thumb. A fixation cross appeared on the screen for 400 ms followed by a 100 ms blank screen in order to orient participants' gazes to the centre of the screen where words were presented in rapid serial visual presentation (RSVP). The words in the sentence were presented for 400 ms with an ISI of 100 ms, with the exception that the 'reinstatement' always appeared with 'the' (e.g., 'The stool'). Participants were provided with eight breaks during the experiment to ensure that they were not getting drowsy. Six practice sentences were viewed in order to familiarise participants with the pace of word presentation. Setup took approximately 20 minutes and the entire experiment took 15 minutes.

Recording procedure

Participants were seated in a comfortable chair in a sound-attenuated darkened room. An elastic cap (Electro-Cap International) with 29 active tin electrodes was placed on the participant's head. The electrodes were located in the standard International 10–20 system locations as well as at additional sites over the left and right hemispheres (see Figure 1). Electrode locations consisted of five sites along the midline (FPz, Fz, Cz, Pz, Oz), three medial electrode sites over each hemisphere (FC1/FC2, C3/C4, CP1/CP2), four lateral electrodes over each hemisphere (F3/F4, FC5/FC6, CP5/CP6, P3/P4), and five peripheral sites over each hemisphere (FP1/FP2, F7/F8, T3/T4, T5/T6, O1/O2). In order to monitor vertical eye movements/blinks and horizontal eye movements, electrodes were placed below the left eye and lateral to the right eye respectively. Electrodes were referenced to the left mastoid and an electrode placed on the right mastoid monitored differential mastoid activity.

The electroencephalogram (EEG) was amplified by a SA Bioamplifier using a bandpass of 0.01 to 40 Hz and was continuously sampled at a rate of 200 Hz. Electrode impedances were kept below 10 k Ω for the eyes and below 5 k Ω at all other sites. For each participant, separate ERPs were averaged off-line at each electrode site for each experimental condition. Trials contaminated with eye artifact or amplifier blockage were not included.

ERP data analysis

All analyses were conducted on mean amplitude values using the 100 ms of activity that preceded word onset as a baseline. Two time windows were chosen for examination: 0–300 ms and 300–500 ms after word onset. The early time window (0–300 ms) was examined to ensure that deviations in later components (i.e., the N400) could not be accounted for by earlier differences between the conditions. The later time window (300–500 ms) was chosen because it corresponds to the time window for the N400. Repeated

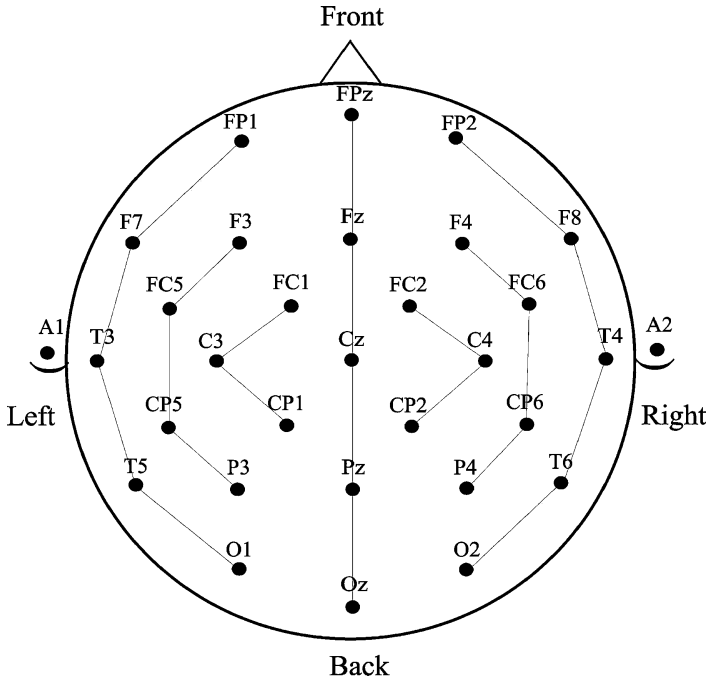


Figure 1. Electrode Montage; the dotted lines represent the four columns used in analyses (i.e., midline, medial, lateral, peripheral).

measures analyses of variance (ANOVAs) were performed on the midline, medial, lateral, and peripheral sites described above. In addition, in all analyses ‘Reinstatement’ (correct, incorrect, control) was entered as a within-subjects factor and, for the medial, lateral, and peripheral analyses; Hemisphere (left, right) was an additional within-subjects factor. A Greenhouse–Geisser correction was applied to all analyses with more than one degree of freedom in the numerator (Greenhouse & Geisser, 1959). In these cases, we report the original degrees of freedom with the corrected p value. These analyses were time-locked to the ‘reinstatement’ of the antecedent in the fifth sentence as well as to the sentence-final word. Significant interactions were explored with simple effects tests. For analyses in which none of the four ANOVAs (i.e., midline, medial, lateral, peripheral) yielded any significant results, we report maximum F -values and minimum p -values.

Results

Artifact contamination from eye movement or amplifier blocking led to the rejection of 4.4% of the trials at the reinstatement of the antecedent

and 6.1% of the trials at the sentence-final word. The number of rejected items did not differ by condition ($p > .05$ for all pairwise comparisons).

Figure 2a (ERP waves to the ‘reinstatement’ word) and 2b (ERP waves to the sentence-final word) show a negative wave component occurring from 100–200 ms and a positive component from 200–300 ms following word onset (the N1–P2 component). The N1–P2 complex was followed by a negative-going component (i.e., the N400) between 300–500 ms.

‘Reinstatement’ word

Early components (0–300 ms). As indicated in Figure 2a, there did not appear to be differences between waveforms to the three ‘Reinstatement’ words in this early time-window. At the ‘reinstatement’ of the antecedent, there was no difference between the ‘Reinstatement’ conditions between 0–300 ms ($F_{\max} < 1$, $p_{\min} = .50$).

N400 (300–500 ms). The three ‘Reinstatement’ conditions did not differ in the N400 time window as indicated in Figure 2a. There were no main effects of ‘Reinstatement’ at any location ($F_{\max} = 2.32$, $p_{\min} = .13$) and ‘Reinstatement’ did not interact with any factor ($F_{\max} = 2.59$, $p_{\min} = .11$).

Sentence-final word

Early components (0–300 ms). At the sentence-final word (see Figure 2b), there were no main effects of ‘Reinstatement’ ($F_{\max} = 2.39$, $p_{\min} = .12$), ‘Reinstatement’ did not interact with Electrode site ($F_{\max} = 1.10$, $p_{\min} = .35$) or Hemisphere ($F_{\max} < 1$, $p_{\min} = .81$), and there were no three-way interactions ($F_{\max} < 1$, $p_{\min} = .21$).

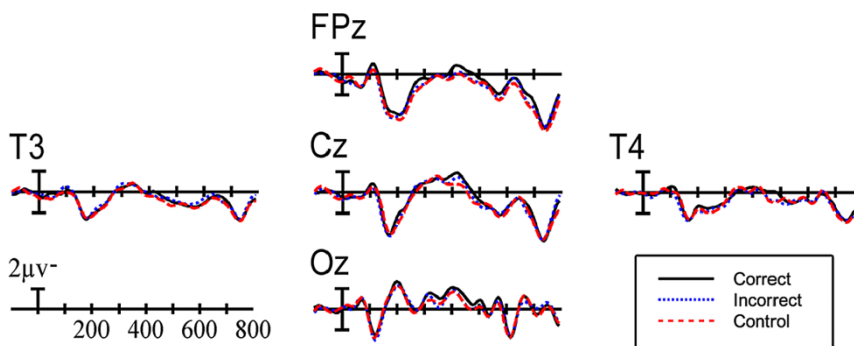


Figure 2a. Wave components at five electrode sites to the ‘reinstatement’ word in Experiment 1.

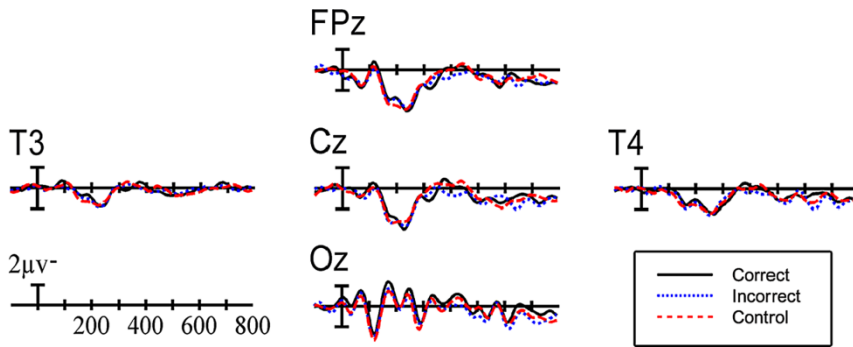


Figure 2b. Wave components at five electrode sites to the sentence-final word in Experiment 1.

N400 (300–500 ms). Again, there were no differences between the three ‘Reinstatement’ conditions (see Figure 2b), reflected by the absence of main effects of ‘Reinstatement’ at any location ($F_{\max} = 3.30$, $p_{\min} = .06$) or interactions between ‘Reinstatement’ and any other factors ($F_{\max} = 1.71$, $p_{\min} = .16$).

Discussion

Experiment 1a verified that participants consistently identified two potential antecedents (e.g., couch, stool) as exemplars of the categorical anaphor (e.g., seat). Experiment 1b demonstrated that without the context (i.e., four preceding sentences) there are no differences between the three ‘Reinstatement’ conditions at the reinstatement of the antecedent or at the sentence final word. Thus, we can attribute any observed differences between the Reinstatement conditions in the following experiment to the process of anaphor resolution rather than item differences. Experiment 2 was conducted to examine anaphor resolution by including the full discourse context.

EXPERIMENT 2

Design and predictions of the current study

ERPs were measured while participants silently read a passage consisting of five sentences. As an index of semantic integration difficulty during anaphor resolution, we measured the N400 to the reinstated antecedent in the fifth sentence.

Our first aim was to explore the time course of anaphor resolution. As noted above, some studies suggest that anaphor resolution occurs online (i.e.,

on the anaphor) whereas other studies have suggested that the resolution process is not complete until clause boundaries in a phrase following an anaphor or, in some cases, at clause boundaries following a reinstated antecedent (Wiley et al., 2001). To evaluate these different possibilities, ERPs to the reinstatement of the antecedent in the final sentence as well as the sentence-final word were inspected. The antecedent was reinstated immediately following the anaphor (with only 100 ms in-between the offset of the anaphor and the onset of the presentation of the reinstatement). Thus, the ease of semantically integrating the reinstated antecedent was taken as an indication that the anaphor had been correctly resolved shortly after anaphor presentation. Alternatively, participants might wait until the sentence-final word of the following clause to resolve the anaphor as observed in previous studies (Wiley et al., 2001), in which case differences between the three conditions would not be observed until the sentence-final word.

It is important to note that the reinstatement of the antecedent, rather than the anaphor, was taken as the earliest point to examine anaphor resolution in the present experiment. Differences were not expected on the anaphor because all three reinstatement conditions were identical at this point (i.e., all of the anaphors had two possible lexico-semantically related antecedents). In other words, at this point in all three conditions, all participants had read the same information (e.g., A stool is found in a bar. A couch is found in a dining room. A clock is found on a wall. At the bar, Henry sat on the seat). Thus, we expected that all participants were resolving the anaphor 'seat' in 'At the bar, Henry sat on the seat' as 'stool' and that this would result in N400 amplitude differences when participants were presented with later information that was consistent (i.e., correct reinstatement) or inconsistent (i.e., incorrect or control reinstatements) with the resolution process. In sum, differences between the three reinstatement conditions were only expected at the reinstatement of the antecedent or later.

Our second aim was to determine how discourse-level and lexico-semantic information interact to correctly resolve a semantically ambiguous anaphor at the point of reinstatement of the antecedent. Several predictions were explored. If the local lexico-semantic relationship between the anaphor and the reinstated antecedent over-rode the global contextual discourse, then one would expect to see the largest N400 to the control antecedent reinstatement but no differences in the N400 amplitude between the correct and incorrect antecedent reinstatement that were both semantically related to the preceding word, the anaphor. For example, after reading 'At the bar, Henry sat on the seat', both 'stool' (the correct reinstatement) and 'couch' (the incorrect reinstatement) – both exemplars of 'seat' – should produce an N400 of the same amplitude, but of smaller magnitude than the N400 to the control antecedent ('clock').

It is also possible that the global discourse context takes precedence over the local lexico-semantic relationship between the anaphor and the antecedent reinstatement. Indeed, Gordon and colleagues (Gordon & Searce, 1995; Kennison & Gordon, 1997; Swaab et al., 2004) have demonstrated that discourse factors (the prominence of an antecedent in discourse) can override lexico-semantic factors (lexical repetition effects) in the resolution of repeated NP anaphors. This was observed when lexical repetitions were separated by several words. In the current experiment, if discourse influences over-ride immediate lexico-semantic associations (i.e., the critical word is preceded by a lexico-semantically related word), we would observe the smallest N400 amplitude to the correct reinstatement ('stool'). The N400 amplitude elicited by incorrect ('couch') and control ('clock') reinstatements would either be the same or perhaps slightly larger to the incorrect than control reinstatement. This latter prediction is based on the behavioural finding that readers are slower to respond to an incorrect probe compared to a control probe following an anaphor (e.g., Lucas et al., 1990).

Finally, discourse and lexico-semantic information may interact such that one would observe the smallest N400 to the correct reinstatement, followed by a medium-sized N400 to the incorrect reinstatement, and the largest N400 to the control. These findings would suggest that there is rapid integration of multiple types of information (e.g., Hagoort, Hald, Bastiaansen, & Petersson, 2004; Van Petten, 1993; Federmeier & Kutas, 1999) and would demonstrate that these multiple sources of information are used effectively during the resolution of anaphors.

Method

Participants

Sixteen participants (mean age = 21.25 years, seven male and nine female) who did not participate in Experiment 1 but who met the same requirements as in Experiment 1, participated in this experiment for monetary compensation. Written informed consent was obtained from all participants in accordance with the guidelines of the Tufts Human Subjects Research Committee.

Procedure

Participants sat approximately 76 cm away from a 19" computer monitor. Participants were instructed to read all sentences for comprehension. Each trial began with the word 'READY' in capital letters in the centre of the screen. When participants were ready to begin the trial, they pressed a button on a game-pad with their left thumb. A fixation cross appeared on the screen for 400 ms followed by a 100 ms blank screen. The first three sentences of

each scenario were then presented self-paced, sentence-by-sentence. Participants advanced to the next sentence by pressing the same button with their left thumb on a game-pad.

Following advancement from the third sentence, a fixation cross again appeared on the screen for 400 ms followed by a 100 ms blank screen to re-orient participants' gazes to the centre of the screen where the rest of the words were presented in RSVP. The fourth sentence was presented word by word, with a stimulus onset asynchrony (SOA) of 400 ms (300 ms word duration and an interstimulus interval (ISI) of 100 ms). The words in the fifth sentence were presented for 400 ms with a 100 ms ISI,¹ with the exception that, as in Experiment 1, the reinstatement always appeared with the article (e.g., 'The stool') in order to minimise the amount of time between the anaphor and the reinstatement. The difference in presentation rate between sentences 4 and 5 was in order to minimise the probability that participants would blink during sentences 4 and 5 and to maximise our ability to examine the N400 to words in sentence 5 (see Figure 3 for an illustration of a typical trial).

In order to ensure that participants were resolving the anaphor, 700 ms after the presentation of the final word of the fifth sentence, a probe word appeared in the centre of the screen in yellow capital letters. This probe word was always the reinstatement that appeared in the fifth sentence. For example, if the fifth sentence was 'The stool was recently purchased' then the probe word was 'stool'. Participants were informed that their task was to decide as quickly and accurately as possible whether the probe word had been referred to in the *fourth* sentence. So, for example, if the fourth sentence was 'At the bar, Henry sat on the seat' and the probe word was 'stool', then the correct response was 'yes' because seat refers to stool. However, if the probe word was 'couch' or 'clock,' the correct response was 'no'. Responses to the probe were made using left and right thumbs corresponding to 'yes' and 'no' on a game-pad. These responses were counterbalanced so that half of the participants responded 'yes' with their left thumb and the other half of the participants responded 'no' with their left thumb. To differentiate them from antecedents and reinstatements, probe words will be referred to as correct probes, incorrect probes, and control probes.

Participants were practiced on six scenarios to ensure that they understood instructions. All participants were able to perform the task following the practice. There were eight breaks during the experiment. The full

¹ This experiment was piloted using 400 ms for both sentences 4 and 5 and found that many trials were lost to blinking. On a post-experimental questionnaire, participants never reported noticing a difference between the presentation rate of sentences 4 and 5.

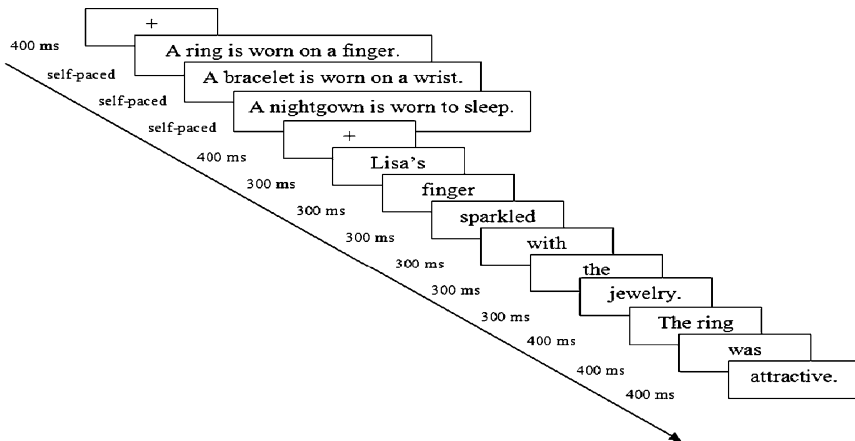


Figure 3. A typical experimental trial in Experiment 1. The time sentences and words were on the computer screen is referenced on the left; the ISI was kept constant at 100 ms.

experimental session took approximately 1.5 hours (35 minutes for setup and 45 minutes for the experiment).

ERP data analysis

All analyses were identical to those performed in Experiment 1 with one exception: in addition to time-locking to reinstatement and sentence-final words, ERPs were also time-locked to the presentation of the anaphor in the fourth sentence.

Results

Behavioural data

We examined participants' responses to probe words by analysing per cent accuracy and RT (in ms) to respond to probe words. Participants were extremely accurate at responding to the probe when it was the correct ($M = 94.64\%$, $SD = 4.80$), control ($M = 99.11\%$, $SD = 1.47$), or incorrect ($M = 92.84\%$, $SD = 3.49$) probe. A one-way repeated measures ANOVA was performed on the accuracy data, revealing a main effect of Probe, $F(2, 30) = 19.54$, $p < .001$. Planned comparisons revealed that participants performed best when presented with the control as the probe (control/correct, $p < .01$; control/incorrect, $p < .001$).

Participants responded to the correct ($M = 642.93$ ms, $SD = 190.87$) probes faster than to the control ($M = 755.48$ ms, $SD = 230.71$) and the incorrect ($M = 767.56$ ms, $SD = 242.06$) probes. This was confirmed by a

one-way repeated measures ANOVA on the log-transformed RT data, $F(2, 30) = 11.04$, $p < .001$. Planned comparisons revealed that participants were significantly faster to respond to the correct probe compared with the control and incorrect probes ($p < .01$ for all pairwise comparisons).

ERP data

All analyses were performed only on trials for which participants responded correctly to the probe.² Artifact contamination from eye movement or amplifier blocking led to the rejection of 4.8% of the trials (to which the probe was responded correctly) at the reinstatement of the antecedent, 6.3% of the trials at the sentence-final word, and 4.3% of the trials at the anaphor. There were no differences between the numbers of rejected trials across the different Reinstatement conditions ($p > .05$ for all pairwise comparisons). Figure 4a (ERP waves to reinstatements) and 4c (ERP waves to sentence-final words) show a negative wave component occurring from 100–200 ms and a positive component from 200–300 ms following word onset (the N1–P2 complex), followed by a negative wave from 300–500 ms after word onset (the N400). In addition, Figure 4b depicts the scalp distribution of each comparison (i.e., control – incorrect reinstatements, incorrect – correct reinstatements, control – correct reinstatements) at 400 ms following onset of the reinstatement word.

Reinstatement of the antecedent

Early components (0–300 ms). At the reinstatement of the antecedent (see Figure 4a), there were no differences between the Reinstatement conditions ($F_{\max} = 1.55$, $p_{\min} = .23$). Reinstatement did not interact with Electrode site ($F_{\max} < 1$, $p_{\min} = .49$) or with Hemisphere ($F_{\max} < 1$, $p_{\min} = .52$).³

N400 (300–500 ms). The N400s evoked by the three reinstatement conditions differed significantly from each other as reflected by main effects of Reinstatement: midline: $F(2, 30) = 6.28$, $p < .01$; medial: $F(2, 30) = 6.27$,

² These analyses were also performed on all trials and yielded similar results.

³ Although the three-way interaction was significant in one column: medial: $F(4, 60) < 1$, $p = .65$; lateral: $F(6, 90) < 1$, $p = .65$; peripheral: $F(8, 120) = 2.89$, $p < .05$, further exploration did not reveal any differences between the three Reinstatement conditions. To further explore the interaction at peripheral sites, separate Reinstatement \times Electrode site ANOVAs were conducted for each Hemisphere. There was no Reinstatement \times Electrode site interaction on the left hemisphere: $F(8, 120) < 1$, $p = .61$, nor on the right hemisphere: $F(8, 120) < 1$, $p = .53$. In addition, these simple effects tests did not show any main effect of Reinstatement at either of these hemispheres, $F_{\max} = 1.28$, $p_{\min} = .29$. Examining each electrode site separately, there were no Reinstatement \times Hemisphere interactions, $F_{\max} = 2.93$, $p_{\min} = .08$.

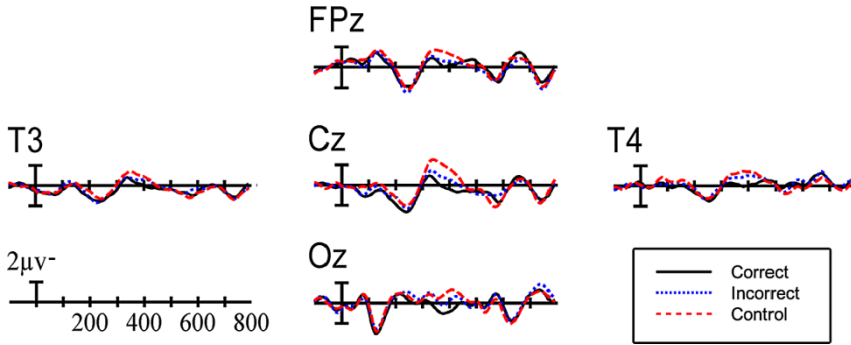


Figure 4a. Wave components at five electrode sites to the reinstatement word in Experiment 2.

$p < .01$; lateral: $F(2, 30) = 6.46, p < .01$; peripheral: $F(2, 30) = 5.76, p < .05$. As seen in Figure 4a and reflected by follow-up t -tests, the correct reinstatements elicited the smallest N400 (midline: correct/incorrect: $p < .10$, correct/control, $p < .01$, incorrect/control: $p < .05$; medial: correct/incorrect: $p = .11$, correct/control: $p < .01$, incorrect/control: $p = .05$; lateral: correct/incorrect: $p < .10$, correct/control: $p < .01$, incorrect/control: $p < .10$; peripheral: correct/incorrect: $p < .05$, correct/control: $p = .05$, incorrect/control: $p < .10$).

In addition, Reinstatement \times Electrode site interactions were observed at all electrode columns: midline: $F(8, 120) = 2.63, p < .10$; medial: $F(4, 60) = 3.30, p < .05$; lateral: $F(6, 90) = 2.85, p < .05$; peripheral: $F(8, 120) = 2.62, p < .10$. Examination of each electrode site separately showed that the effect of Reinstatement was most pronounced at centroparietal sites (as depicted in Figure 4b). All three conditions differed statistically at centroparietal sites ($p < .05$) whereas at more anterior sites, the correct/incorrect comparison was only marginally statistically significant ($p < .10$) and at more posterior sites, the incorrect/control comparison was only marginally statistically significant ($p < .10$).

The analyses also revealed three-way interactions: medial: $F(4, 60) = 5.55, p < .01$; lateral: $F(6, 90) = 2.15, p < .10$; peripheral: $F(8, 120) = 2.12, p < .10$. To further explore these interactions, separate ANOVAs (Reinstatement \times Hemisphere) were conducted at each Electrode site. At anterior sites, a Reinstatement \times Hemisphere interaction was observed: medial (FC1/FC2): $F(2, 30) = 3.68, p = .05$; lateral (FC5/FC6): $F(2, 30) = 2.64, p < .10$; peripheral (FP1/FP2): $F(2, 30) = 4.11, p < .05$, reflecting a larger difference between Reinstatement conditions in the right hemisphere than in the left hemisphere. At posterior sites, no Reinstatement \times Hemisphere interactions were found ($F_{\max} = 1.74, p_{\min} = .19$) as the main effect of Reinstatement was similar on both hemispheres.

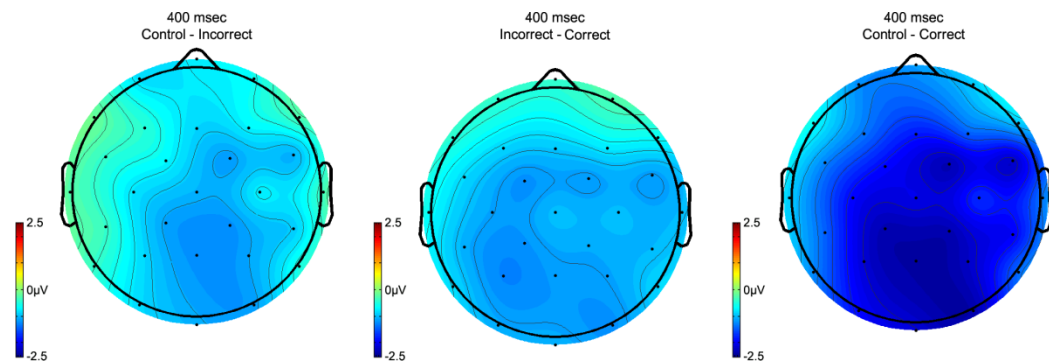


Figure 4b. Scalp distribution of comparisons between control – incorrect reinstatements, incorrect – correct reinstatements, and control – correct reinstatements at 400 ms following onset of the reinstatement word in Experiment 2.

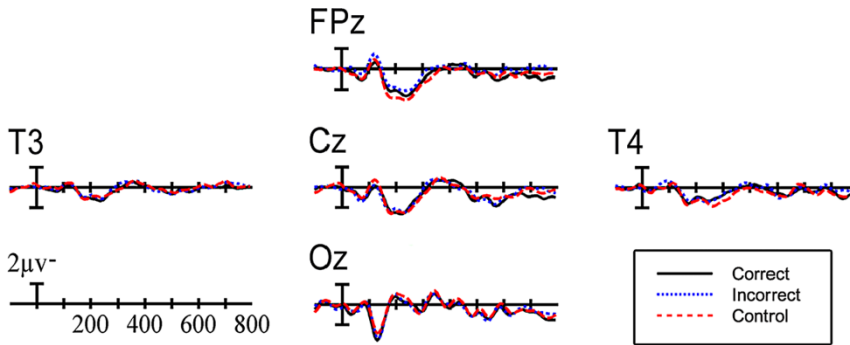


Figure 4c. Wave components at five electrode sites to the sentence-final word in Experiment 2.

Sentence-final word

Early components (0–300 ms). At the sentence-final word, there was no main effect of Reinstatement ($F_{\max} < 1$, $p_{\min} = .43$). Reinstatement did not interact with Hemisphere ($F_{\max} = 1.12$, $p_{\min} = .32$) but, at peripheral sites, there was an interaction with Electrode site: midline: $F(8, 120) = 2.68$, $p < .10$; medial: $F(4, 60) = 2.22$, $p = .13$; lateral: $F(6, 90) = 2.21$, $p = .11$, peripheral: $F(8, 120) = 4.73$, $p < .05$. There was no three-way interaction ($F_{\max} = 1.04$, $p_{\min} = .38$).⁴

N400 (300–500 ms). At the final word of the sentence, there were no main effects of Reinstatement ($F_{\max} < 1$, $p_{\min} = .47$). In addition, there were no interactions with Reinstatement and any other factor ($F_{\max} = 3.47$, $p_{\min} = .06$). No other effects reached statistical significance.

The N400 to the anaphor

In order to ensure that differences between the three Reinstatement conditions were not observed prior to the reinstatement of the antecedent, N400 amplitude to the anaphor was also examined. As discussed in the Introduction, differences were not expected because all three reinstatement conditions were identical at this point (i.e., all of the anaphors had two possible lexico-semantically related antecedents) and all participants had read the same information. Earlier differences (i.e., at the anaphor) between the three reinstatement conditions would limit interpretations of the results with respect to the reinstatement (this will be addressed in greater detail in the General Discussion). A visual inspection of Figure 4d suggested that

⁴ Although early differences (0–300 ms) were observed to the sentence-final word at peripheral sites, these differences cannot account for the lack of effects seen in the N400 window.

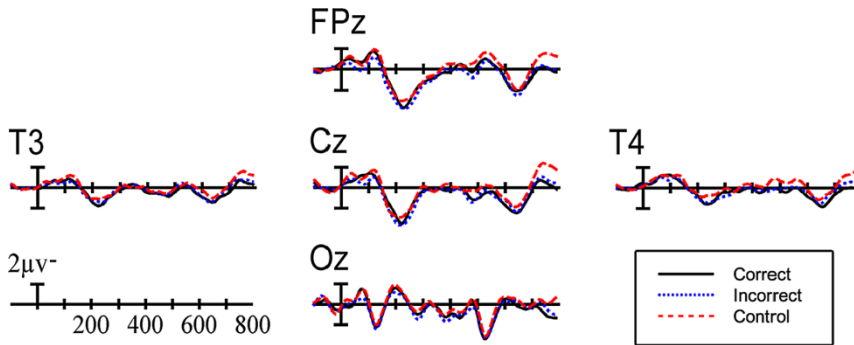


Figure 4d. Wave components at five electrode sites to the anaphor in Experiment 2.

there were, indeed, no N400 amplitude differences at the anaphor, as reflected by the absence of main effects ($F_{\max} = 1.54$, $p_{\min} = .23$), interactions with Electrode site ($F_{\max} = 1.19$, $p_{\min} = .32$) or Hemisphere ($F_{\max} = 1.42$, $p_{\min} = .26$), or three-way interaction ($F_{\max} = 1.18$, $p_{\min} = .32$).

Discussion

In this experiment, participants read either a correct (e.g., 'stool'), incorrect (e.g., 'couch'), or control (e.g., 'clock') reinstatement of an antecedent following an ambiguous anaphor (e.g., 'seat') in a discourse context. The results demonstrate that, immediately following an anaphor, correct reinstatements are the most easily semantically integrated into their preceding contexts, as evidenced by the smallest N400 evoked by these words. Control reinstatements (e.g., 'clock') elicited the largest N400, reflecting the difficulty readers had integrating these words into the discourse context. Incorrect reinstatements (e.g., 'couch') evoked intermediate N400s. This suggests that they were influenced both by their local semantic relationship with their preceding anaphor and by the global discourse context. As we expected, no differences between the three conditions were found at the sentence ending or on the anaphor. The significance of this latter point will be considered further in the General Discussion.

Although differences in the N400 were observed between the three reinstatement conditions, we cannot definitively attribute these differences to factors associated with anaphor resolution. An alternative possibility is that participants did not treat the items presented in the first three sentences as potential antecedents. In other words, it may be the case that due to the nature of the stimuli (i.e., the first three sentences were generic and, in some but not all cases, introduced pragmatic information) participants did not need to use the information in the first three sentences to inform their resolution of the anaphor in the fourth sentence. Thus, participants may have

ignored the first three sentences and merely attended to sentences four and five. Similarly, because of the generic nature of the first three sentences, participants may not have treated these sentences as part of the discourse and may therefore not have attempted to integrate information across the full five-sentence discourse. For example, in the fourth sentence, 'At the bar, Henry sat on the seat', participants may have realised that seat likely refers to a stool and the full discourse context (the three preceding sentences) may have been unnecessary for comprehension.

This leads to an alternative interpretation of the ERP results with respect to the reinstatement: the reinstatements (e.g., the stool) may have been regarded as possible anaphors (rather than reinstatements) to the immediately preceding reference. In other words, readers may have treated the anaphor in the fourth sentence (e.g., seat) as the antecedent. In turn, the reinstatement in the fifth sentence may have been treated as the anaphor. This would lead to the same pattern of N400s observed in Experiment 2. However, it would not inform us about the resolution process, if any, for the original anaphor. To rule out this interpretation, it is necessary to examine probe responses after reading the final two sentences (i.e., sentences 4–5) without the preceding context (i.e., sentences 1–3). If readers did ignore the first three sentences and, in turn, treat the anaphor in the fourth sentence as the antecedent, then one would expect to replicate the results of Experiment 2 even when the first three sentences are omitted. However, if readers did treat the items presented in the first three sentences as potential antecedents, then one would expect a different pattern of results when the first three sentences are omitted.

EXPERIMENT 3

Experiment 3 was conducted in order to examine whether differences in ERPs evoked to the antecedent reinstatements observed in Experiment 2 were due to the integration of information across the full five-sentence discourse (i.e., due to the reinstatement of the antecedent) or only due to the integration over the final two sentences of the scenarios. ERPs were measured as subjects read the final two sentences without the preceding context.

We had interpreted the results of Experiment 2 as suggesting that participants had used the context provided in sentences 1–3 to constrain the interpretation of the ambiguous categorical anaphor at the end of the fourth sentence. By omitting these sentences, we predicted that sentences with correct 'reinstatements' would still be comprehensible and sentences with control 'reinstatements' would remain incomprehensible. However, we also predicted that participants would encounter a problem evaluating the

incorrect 'reinstatements'. To illustrate why this would be the case, compare the following scenarios from the current Experiment 3 with a scenario from Experiment 2:

From the current experiment (final two sentences only): At the bar, Henry sat on the seat. The couch was recently purchased.

From Experiment 2 (all five sentences): A couch is found in a living room. A clock is found on a wall. A stool is found in a bar. At the bar, Henry sat on the seat. The couch was recently purchased.

In the two-sentence scenario, although not entirely pragmatic, it would be reasonable to sit on a couch at a bar. Thus, the 'seat' referred to in the preceding sentence could be interpreted as 'couch'. However, in scenario 2, given the constraints imposed by the first three sentences, it would be incorrect to interpret 'seat' in sentence 4 as a 'couch'. Corroborating this point, in Experiment 1, there were no participants who erroneously responded 'yes' to the incorrect probe 'couch'. By omitting the constraint provided by the first three sentences of the discourse, we predicted that participants would find it difficult to determine whether incorrect 'reinstatements' are, in fact, antecedents of the 'anaphor'. We predicted that this difficulty would manifest as a large percentage of inaccurate responses and slow response times to incorrect probes, as well as by differential electrophysiological activity to incorrect reinstatements with correct responses (i.e., 'no') compared with incorrect responses (i.e., 'yes') to probes. Specifically, we predicted that incorrect reinstatements with incorrect responses would pattern with correct reinstatements with accurate responses to probes (i.e., 'yes' responses), evoking smaller amplitude N400s, whereas incorrect reinstatements with incorrect responses would pattern with control reinstatements with correct responses (i.e., 'no'), evoking larger amplitude N400s.

Method

Participants

Sixteen native-English speakers with a mean age of 19.06 years (seven male and nine female) meeting the same inclusion criteria as those in Experiment 1 participated for monetary compensation after giving informed consent. None of the participants had participated in the previous experiments.

Procedure

The procedure was identical to that of Experiment 2 with the exception that participants did not read the first three sentences.

ERP data analysis

Based on our predictions and confirmed by the behavioural results (see below), we expected four distinct groups of items to emerge: correct 'reinstatements' with correct responses, control 'reinstatements' with correct responses, 'incorrect reinstatements' with correct responses (referred to as incorrect 'no'), and 'incorrect reinstatements' with incorrect responses (referred to as incorrect 'yes'). Thus, all four of these 'Reinstatement' conditions were entered into ANOVAs as a within-subjects factor. All other analyses were conducted in the same way as in the previous two experiments.

Results

Behavioural data

As in Experiment 2, we examined participants' responses to probe words by analysing per cent accuracy and RT (in ms) to respond to probe words. As predicted, participants were extremely accurate at responding to the probe word when it was the correct ($M = 88.64\%$, $SD = 16.06$) or control ($M = 92.06\%$, $SD = 16.45$) probe. However, unlike in Experiment 2, participants were extremely inaccurate in their responses to the incorrect ($M = 55.46\%$, $SD = 15.11$) probe. A one-way repeated measures ANOVA was performed on the accuracy data, revealing a main effect of Probe, $F(2, 30) = 68.93$, $p < .001$. Planned comparisons confirmed that, whereas participants' performance to correct and control probes was the same ($p > .10$), accuracy was worse when participants were presented with the incorrect probe (incorrect/correct, $p < .001$; incorrect/control, $p < .001$).

In addition, participants responded to the incorrect ($M = 1263.27$ ms, $SD = 614.85$) probes slower than to the control ($M = 1020.40$ ms, $SD = 395.35$) and the correct ($M = 907.41$ ms, $SD = 354.55$) probes. This was confirmed by a one-way repeated measures ANOVA on the log-transformed RT data, $F(2, 30) = 14.20$, $p < .001$. Planned comparisons revealed that all three conditions were significantly different. Specifically, participants were slowest at responding to the incorrect probe (incorrect/correct, $p < .01$, incorrect/control, $p < .01$) and fastest at responding to the correct probe compared with the control and incorrect probes (correct/control, $p < .05$).

ERP data

Figure 5a (ERP waves to 'reinstatement' word) and Figure 5b (ERP waves to sentence-final word) show a negative wave component occurring from 100–200 ms and a positive component from 200–300 ms following word onset (the N1–P2 component). The N1–P2 complex was followed by a negative-going component (i.e., the N400) between 300–500 ms. Artifact

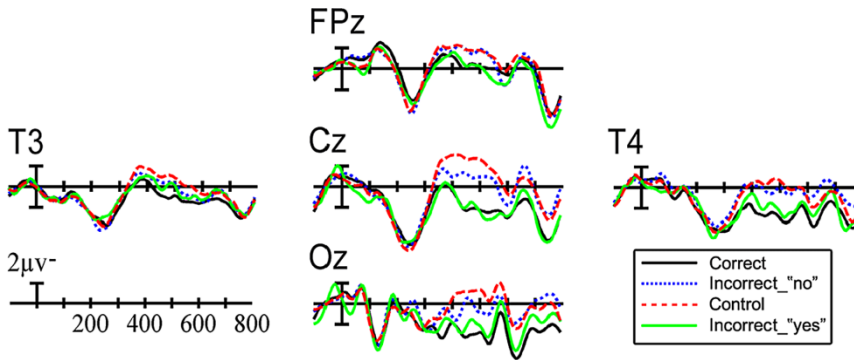


Figure 5a. Wave components at five electrode sites to the 'reinstatement' word in Experiment 3.

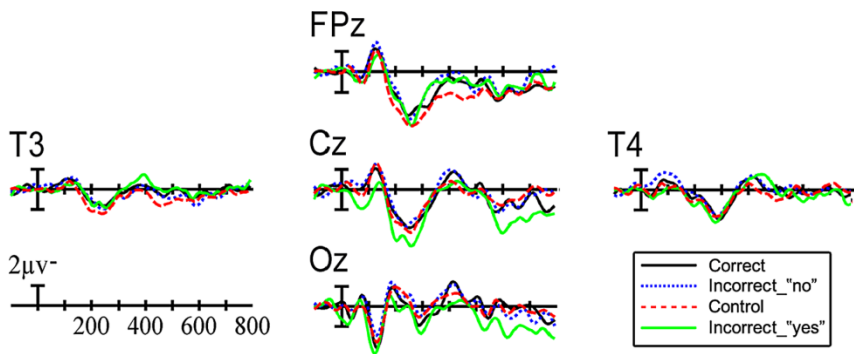


Figure 5b. Wave components at five electrode sites to the sentence-final word in Experiment 3.

contamination from eye movement and amplifier blocking led to the rejection of 11.5% of the trials at the reinstatement of the antecedent and 12.9% of the trials at the sentence-final word. The number of rejected items did not differ by condition ($p > .05$ for all pairwise comparisons). For analyses in which none of the four ANOVAs (i.e., midline, medial, lateral, peripheral) yielded any significant results, we report maximum F -values and minimum p -values.

'Reinstatement' of the antecedent

Early components (0–300 ms). At the 'reinstatement' of the antecedent (see Figure 5a), there were no differences between the 'Reinstatement' conditions ($F_{\max} < 1$, $p_{\min} = .68$). 'Reinstatement' did not interact with Electrode site ($F_{\max} < 1$, $p_{\min} = .50$) or with Hemisphere ($F_{\max} < 1$, $p_{\min} = .63$). There was no three-way interaction ($F_{\max} = 1.95$, $p_{\min} = .11$).

N400 (300–500 ms). The N400s evoked by the three ‘reinstatement’ conditions differed significantly from each other as reflected by a main effect of ‘Reinstatement’: midline: $F(3, 45) = 7.72, p < .001$; medial: $F(3, 45) = 8.22, p < .001$; lateral: $F(3, 45) = 8.05, p < .001$; peripheral: $F(3, 45) = 9.22, p < .001$. As seen in Figure 5a and reflected by follow-up *t*-tests, the incorrect ‘yes’ ‘reinstatement’ condition elicited an N400 similar to the correct ‘reinstatement’ condition ($p > .10$ for midline, medial, lateral, and peripheral comparisons) whereas the incorrect ‘no’ and control ‘reinstatement’ conditions evoked similar amplitude N400s ($p > .05$ for midline and medial comparisons, $p > .10$ for lateral and peripheral comparisons). Specifically, incorrect ‘yes’ and correct ‘reinstatements’ elicited smaller amplitude N400s than incorrect ‘no’ and control ‘reinstatements’ (for midline, medial, lateral, and peripheral comparisons: incorrect ‘yes’/control: $p < .01$; incorrect ‘yes’/incorrect ‘no’: $p < .05$; correct/incorrect ‘no’: $p < .05$; correct/control: $p < .05$).

Sentence-final word

Early components (0–300 ms). At the sentence-final word, there was an effect of ‘Reinstatement’ at midline and medial sites: midline: $F(3, 45) = 3.26, p < .05$; medial: $F(3, 45) = 4.01, p < .05$; lateral: $F(3, 45) = 2.67, p > .05$; peripheral: $F(3, 45) = 1.69, p > .10$. The incorrect ‘yes’ condition elicited the most positive response at both midline and medial sites. ‘Reinstatement’ did not interact with Hemisphere ($F_{\max} = 1.81, p_{\min} = .18$) or Electrode site ($F_{\max} = 1.86, p_{\min} = .12$). There was no three-way interaction ($F_{\max} < 1, p_{\min} = .57$).

N400 (300–500 ms). At the final word of the sentence, there was no effect of ‘Reinstatement’ ($F_{\max} = 1.20, p_{\min} = .32$). In addition, ‘Reinstatement’ did not interact with Hemisphere ($F_{\max} < 1, p_{\min} = .53$) or Electrode site ($F_{\max} = 2.21, p_{\min} = .10$) and there was no three-way interaction ($F_{\max} = 1.74, p_{\min} = .14$).

Discussion

In order to examine whether participants in Experiment 2 integrated information across all five sentences of the discourse, the first three sentences of the discourse were omitted. Both the behavioural and ERP data support the conclusion that participants were in fact using information from the first three sentences to resolve the anaphor presented in the fourth sentence. Specifically, when the first three sentences were not present, participants were less certain whether incorrect ‘reinstatements’ were, in fact, incorrect. This was observed in the behavioural data as a large percentage of incorrect ‘yes’

responses to incorrect probes (i.e., approximately 50% accuracy) and longer RTs. Neurally, incorrect 'yes' 'reinstatements' evoked a similar amplitude N400 as correct 'reinstatements' which was smaller than the N400 evoked by incorrect 'no' and control 'reinstatements'. Taken together, these results confirm that readers were integrating information across all five sentences of discourse in Experiment 2.

GENERAL DISCUSSION

The present experiments were designed to examine the interactions between discourse context and lexico-semantic information on anaphor resolution during online language processing. To date, few studies have looked at the time course of anaphor resolution using ERPs and none, to our knowledge, have examined ERP components to a reinstated antecedent. Taken together, the pattern of N400 modulation in Experiments 1, 2, and 3 suggest that after an anaphor was presented, a correct reinstatement was the easiest to semantically integrate into its preceding context, as evidenced by the smallest N400. This was followed by increased difficulty in semantically integrating an incorrect reinstatement, as evidenced by an intermediate N400, and most difficulty in semantically integrating a control reinstatement, as evidenced by the largest N400.⁵ Importantly, these results cannot be explained by sentence-level effects as no differences were observed when the entire preceding discourse context was excluded in Experiment 1b. Additionally, Experiment 3, that excluded the first three sentences, confirmed that these results reflect the integration of information across the full five-sentence discourse.

We think that the most likely explanation for this pattern of findings is that readers quickly integrated multiple types of information within a discourse to resolve anaphors. We first consider the two alternative possibilities discussed in the Introduction: first, that effects were driven purely by lexico-semantic association with no influence of discourse, and second, that the discourse context completely over-rode any lexico-semantic effects.

Because correct reinstatements were presented in contexts in which two words were semantically related to this item, incorrect reinstatements were semantically related to one word, and control reinstatements were not semantically related to any of the words within the preceding context, one could argue that the combination of these lexico-semantic influences led to the observed pattern of findings. For example, following the sentence

⁵ Although small differences were observed shortly after word onset (i.e., 0–300 ms following word onset), neither the magnitude nor pattern of these effects can explain differences in the N400 between the three conditions. These differences will not be discussed further.

'At the bar, Henry sat on the seat', both 'bar' and 'seat' might prime 'stool' (the correct reinstatement) leading to the smallest N400 amplitude; only 'seat' would prime 'couch' (the incorrect reinstatement) leading to the medium-sized N400 amplitude and there would be no priming of 'clock' (the control reinstatement), leading to the largest N400 amplitude. We believe that such effects are unlikely to explain our findings for two reasons. First, although some studies have demonstrated additive activation of multiple semantically related primes, these priming effects do not appear to operate over long distances – they are significantly smaller when even a single word intervenes between the two related words (e.g., Balota & Paul, 1996). In the present study, the mean distance between the disambiguating word (e.g., 'bar') and the anaphor was 3.20 words ($SD = 1.18$); that is, approximately 1250 ms separated the disambiguating word from the anaphor. Second, and most importantly, if our results were driven purely by these lexico-semantic associations, one would also expect to see differential modulation of the N400 across the three conditions at the point of the anaphor itself which also differed across scenario types in terms of its preceding semantic associations. Specifically, the anaphor (e.g., seat) was lexico-semantically related both to its previously presented correct (e.g., stool) and incorrect (e.g., couch) antecedents but not to its control (e.g., clock) antecedent. If processing was driven only by lexico-semantic influences, one would expect to see a larger amplitude N400 to the anaphor in the control condition compared with anaphors in the correct and incorrect conditions. No such N400 modulation was, in fact, observed.

Our data also suggest that discourse influences failed to completely over-ride lexico-semantic factors, as this would have predicted no difference in the N400 between the incorrect and control reinstatements. Of note, this finding differs from that of Swaab, Camblin, and Gordon (2004) who found that lexical repetition effects could be over-ridden by discourse factors (such as focus of attention manipulations) in anaphor resolution. In Swaab et al., however, lexical repetitions were separated by several intervening words. This contrasts with the present study where the reinstatement was immediately preceded by the lexico-semantically related anaphor. It is possible that lexico-semantic associations that immediately precede critical words may interact with discourse-level information online, but that discourse-level factors can over-ride lexical effects with greater distance between these associations or repetitions. Future studies should further explore the relationship between discourse and lexical influences by varying distances between lexico-semantically related items.

In the current study, there are two possible mechanisms by which local semantic and global discourse information may have interacted on the N400

to produce the pattern of findings we observed at anaphor resolution: first through the interaction of discourse context with the lexico-semantic relationships between the critical word and previously presented content words, and second, through the interaction of discourse context with semantic memory structure.

On the first account, both local lexico-semantic relationships and global discourse context together influenced the amplitude of the N400. This would be consistent with findings of Van Petten (1993) who embedded semantically related (i.e., moon-stars) and semantically unrelated (i.e., insurance-refused) word-pairs in congruous (i.e., semantically meaningful) and anomalous (i.e., syntactically correct but semantically meaningless) sentences. The reduction in N400 amplitude between the first and second words in the pairs was indexed with the presumption that the greater the difference, the greater the ease of semantic integration of the second word. When the word-pairs were unrelated, there was an N400 amplitude difference between the first and second words in the pairs only when the sentences were congruous, but not when they were anomalous. However, when the word-pairs were related, the N400 amplitude reduction for the second word in the pairs was significant in both congruous and anomalous sentences, with the greatest reduction in the congruent sentences. Similarly, Hoeks, Stowe, and Doedens (2004) found interactive effects of sentence context and lexico-semantic influences on N400 amplitude. Specifically, when sentence endings were lexico-semantically related to preceding words in a sentence, N400 amplitude was similar in sentences in which this ending resulted in anomalous (in sentences in which the final word was weakly constrained by context) or congruous sentences (in sentences in which the final word was strongly constrained by context). However, when sentence-final words were unrelated to preceding words resulting in anomalous endings in both constraint conditions, sentential context differentially impacted processing of these unrelated words, with the N400 being more sensitive (larger amplitude) to anomalous endings in strongly constrained sentences. Thus, the amplitude of the N400 was influenced by several factors in these studies: first, by the lexico-semantic associations between the embedded words as evidenced by an N400 amplitude reduction to the related words regardless of sentence congruity; second, by the sentential context as evidenced by an N400 reduction to congruous sentences but not to anomalous sentences when the word-pairs were unrelated in the Van Petten study, as well as larger amplitude N400 in strong-constraint sentences relative to weak-constraint sentences with unrelated words in Hoeks et al. (2004); and third, by a combination of both lexico-semantic and sentential factors as evidenced

by a more pronounced N400 amplitude reduction with congruous sentences that also had embedded related word-pairs in the Van Petten study, as well as by differential effects of lexico-semantic associations based on sentential constraint in Hoeks et al. (2004).

We observed a similar pattern of findings in our discourse scenarios. Regardless of discourse context, the amplitude of the N400 was reduced purely by lexico-semantic associations, as demonstrated by a smaller amplitude N400 for incorrect reinstatements compared with control reinstatements. In addition, consistent with Van Petten's (1993) findings, the correct reinstatement condition elicited the smallest N400 as it received a boost from both lexico-semantic associations as well as congruity with the preceding discourse context.

Both Van Petten (1993) and Hoeks et al. (2004) examined the interaction between lexico-semantic effects and context *within* sentences. The present study is unique in that interactions between lexico-semantic effects and context were explored across sentences within whole discourse. Of note, the influences of lexico-semantic relationships on the N400 may have been the result of the relationship between the antecedent and the anaphor and/or the relationship between the anaphor and the critical word (i.e., the reinstatement of the antecedent). In examinations of repeated NP anaphors (e.g., using the anaphor 'Mark' to refer to the antecedent 'Mark'), Gordon and colleagues have demonstrated that results of probe verification tasks (which assume that responses to a probe word – e.g., 'Mark' – presented immediately following an anaphor reflects the ease of the resolution process) may reflect the relationship between the anaphor and the probe rather than the relationship between the anaphor and its antecedent (e.g., Gordon, Hendrick, & Foster, 2000).

On the second account, the amplitude of the N400 was modulated by the overlap in semantic features between the correct (and thus expected) reinstatement and the other two conditions. This explanation is suggested by the results of Federmeier and Kutas (1999) who showed that the amplitude of the N400 varied accordingly with the semantic feature overlap of unexpected words to an expected word in a discourse context. Words that had high overlap (e.g., the word 'roses', referred to as a within-category violation) with an expected word (e.g., 'tulips') elicited an N400 amplitude between that of the expected word and words that had low overlap (e.g., the word 'pines', referred to as a between-category violation). It is possible that, in the current study, the incorrect reinstatement condition acted similarly to the within-category violation condition in the Federmeier and Kutas study. In other words, the incorrect reinstatement, 'couch', may have elicited a smaller N400 than the control condition, 'clock', because it is a member of

the same category (i.e., seat) as the correct reinstatement, 'stool'. Thus, 'couch' has a greater overlap in semantic features with 'stool' than 'clock'. If participants activated the correct antecedent upon reading the anaphor, then, according to Federmeier and Kutas, the semantic features of 'stool' were activated and would result in a pattern of N400 amplitudes similar to the one observed.

Regardless of whether the N400 was modulated by the interaction of discourse context with lexico-semantic relationships between the target word and previous content words, or by the interaction of discourse context with semantic memory structure, its implications are the same: discourse context and lexico-semantic/semantic memory can interact in the service of anaphor resolution. Specifically, correct reinstatements were easier to semantically integrate into their discourse context than incorrect reinstatements, even though both were lexico-semantically related to the anaphor. In other words, lexico-semantic facilitative effects between anaphor and antecedent were not sufficient to correctly resolve the anaphor; discourse context was necessary. This finding is consistent with the results of behavioural studies of anaphor resolution that have demonstrated shorter RT (e.g., Chang, 1980; Dell, McKoon, & Ratcliff, 1983; Gernsbacher, 1989; McKoon & Ratcliff, 1980) and eye gazes (e.g., Garrod, O'Brien, Morris, & Rayner, 1990) to a correct probe following an anaphor. Unlike behavioural studies that have yielded equivocal results concerning the time course for integration (e.g., immediate: Lucas et al., 1990; delayed: Wiley et al., 2001), our findings suggest that a correct antecedent is semantically integrated into the discourse context immediately following a categorical anaphor. In addition, our findings demonstrate that the resolution process was completed prior to the sentence-final word, as evidenced by similar amplitude N400s between the three Reinstatement conditions at this point.

Gernsbacher (1989) postulated that discourse context aids in anaphor resolution by both activating the correct antecedent and suppressing activation for the incorrect antecedent, although, as discussed in the Introduction, the results of behavioural studies supporting this theory have been equivocal (Lucas et al., 1990). Taken alone, the ERP findings in the current study do not clearly distinguish between these mechanisms: the increased boost by the discourse context upon correctly resolving the anaphor may arise either because the biasing context provides easier access to the correct antecedent by way of facilitation for the correct antecedent or by way of inhibition of the incorrect antecedent. On the other hand, the behavioural data do provide some insights into the mechanism by which discourse context aids anaphor resolution. In Experiment 1, we found poorer accuracy (compared with control probes) as well as slower responses (as

compared with correct probes) to incorrect probes.⁶ This suggests that re-processing an incorrect antecedent at some point after anaphor resolution may be more difficult than re-processing either correct antecedent or control probes. These results may reflect inhibitory processes.⁷

Other recent studies have focused on ERPs at the point of presentation of the anaphor itself where they have reported a LAN ERP component. Specifically, Anderson and Holcomb (2005) found that nouns preceded by a definite article (i.e., 'the') elicited a larger LAN compared with nouns that were preceded by an indefinite article (i.e., 'a'). In addition, Van Berkum et al. (1999a) found that ambiguous anaphors (i.e., anaphors without unique referents) elicited larger LANs compared with anaphors with unique referents. These findings have been interpreted as evidence that people immediately attempt to locate a correct antecedent for an anaphor and that this process may entail a working memory load. In the current study, we focused on ERP components to reinstated antecedents rather than the anaphors themselves where we did not predict any differences between conditions: unlike Anderson & Holcomb (2005), all nouns used in the comparison were preceded by a definite article (particularly relevant for comparisons to Anderson and Holcomb) and, unlike Van Berkum et al. (1999a), all anaphors were lexico-semantically ambiguous (e.g., without the preceding context, 'seat' could refer to 'stool' or 'couch'). Our findings are consistent with and build upon these previous findings. They confirm that the comprehension system attempts to establish cohesion of an anaphor with its antecedent as soon as an anaphor is presented, and they further suggest that this lexico-semantic and discourse information are both used to correctly resolve such anaphors during online processing. In sum, our findings add to a growing body of literature suggesting that readers use different types of information (i.e., discourse context and lexico-semantic or semantic memory features) in processing incoming information and generating a global discourse representation. We have demonstrated that such

⁶ Responses to a correct probe are consistent with a speed-accuracy trade-off in that participants were slightly (but significantly) less accurate at responding to a correct probe compared with a control probe but also significantly faster in their responses. In the same respect, although participants were significantly more accurate at responding to a control probe, they were slower in their responses as compared with the correct condition. Responses to incorrect probes were both significantly less accurate (as compared with a control) and significantly slower than responses to correct probes (they were no different in speed from control probes). Thus, we are focusing on these responses because they do not seem to be affected by any speed-accuracy trade-off.

⁷ However, an alternative interpretation is that these slower reaction times to incorrect probes reflect response competition involved in saying 'no' to a thematically related probe (i.e., stool and couch are related in that they are both types of seats), consistent with McNamara and McDaniel's (2004) account of increased RT for contextually inappropriate meanings for ambiguous homographs. Our results are unable to differentiate between these two accounts.

processes occur at the point of correctly resolving an anaphor to build up discourse cohesion.

Manuscript received September 2005

Revised manuscript received September 2006

First published online April 2007

REFERENCES

- Almor, A. (1999). Noun-phrase anaphors and focus: The informational load hypothesis. *Psychological Review*, *106*, 748–765.
- Anderson, J., & Holcomb, P. J. (2005). An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain and Language*, *94*, 200–216.
- Balota, D. A., & Paul, S. T. (1996). Summation of activation: Evidence from multiple primes that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 827–845.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*, 343–355.
- Chang, F. R. (1980). Active memory processes in visual sentence comprehension: Clause effects and pronominal reference. *Memory and Cognition*, *8*, 58–64.
- Corbett, A. T., & Chang, F. R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory and Cognition*, *11*, 283–294.
- Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior*, *22*, 121–132.
- Duffy, S. A., & Rayner, K. (1990). Eye movements and anaphor resolution: Effects of antecedent typicality and distance. *Language and Speech*, *33*, 103–119.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469–495.
- Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, *50*, 259–281.
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove, UK: Psychology Press.
- Garrod, S. C., O'Brien, E. J., Morris, R. K., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 250–257.
- Garrod, S. C., & Sanford, A. J. (1977). Interpreting anaphoric relations: The integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior*, *16*, 77–90.
- Garrod, S. C., & Sanford, A. J. (1994). Resolving sentences in a discourse context: How discourse representations affect language understanding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 675–698). San Diego: Academic Press, Inc.
- Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, *32*, 99–156.
- Gordon, P. C., Hendrick, R., & Foster, K. L. (2000). Language comprehension and probe-list memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 766–775.
- Gordon, P. C., & Scarce, K. A. (1995). Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory and Cognition*, *23*, 313–323.
- Greene, S. T., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 266–283.

- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 32, 95–112.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669), 438–441.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13, 512–521.
- Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19, 59–73.
- Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, 30, 47–61.
- Kennison, S. M., & Gordon, P. C. (1997). Comprehending referential expressions during reading: Evidence from eye tracking. *Discourse Processes*, 24, 229–252.
- King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 376–395.
- Kucera, H., & Francis, W. H. (1967). *Computation analysis of present-day American English*. Providence, RI: Brown University Press.
- Kuperberg, G. R., Sitnikova, T., Caplan, C., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17, 117–129.
- Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, 8, 533–572.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4, 463–470.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic anomalies. *Nature*, 307, 161–163.
- Levine, W. H., Guzman, A. E., & Klin, C. M. (2000). When anaphor resolution fails. *Journal of Memory and Language*, 43, 594–617.
- Lucas, M. M., Tanenhaus, M. K., & Carlson, G. N. (1990). Levels of representation in the interpretation of anaphoric reference and instrument inference. *Memory and Cognition*, 18, 611–631.
- MacDonald, M. C., & MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language*, 29, 469–492.
- McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structure involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior*, 19, 668–682.
- McNamara, D. S., & McDaniel, M. A. (2004). Suppressing irrelevant information: Knowledge activation or inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 465–482.
- Nicol, J. (1988). *Coreference processing during sentence comprehension*. Unpublished manuscript, Cambridge.
- O'Brien, E. J., Duffy, S. A., & Myers, J. L. (1986). Anaphoric inference during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 346–352.
- Rugg, M. D. (1984). Event-related potentials in phonological matching tasks. *Brain and Language*, 23, 225–240.
- Sanford, A. J., & Garrod, S. C. (1994). Selective processing in text understanding. In M. A. Gernsbacher (Ed.), *Handbook in psycholinguistics* (pp. 699–719). San Diego: Academic Press Inc.
- St George, M., Mannes, S., & Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience*, 6, 70–83.

- Swaab, T. Y., Camblin, C. C., & Gordon, P. C. (2004). Electrophysiological evidence for reversed lexical repetition effects in language processing. *Journal of Cognitive Neuroscience*, *16*, 715–726.
- Van Berkum, J. J. A. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things? In M. Carreiras & C. Jr. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 229–270). New York: Psychology Press.
- Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147–182.
- Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657–671.
- Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400. *Cognitive Brain Research*, *17*, 701–718.
- Van Gompel, R. P. G., & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition*, *90*, 255–264.
- Van Petten, C. (1993). A comparison of lexical and sentence-level context effects and their temporal parameters. *Language and Cognitive Processes*, *8*, 485–532.
- Wiley, J., Mason, R. A., & Myers, J. L. (2001). Accessibility of potential referents following categorical anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1238–1249.