available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400

Lea A. Hald[a,*], *Esther G. Steenbeek-Planting*[b], *Peter Hagoort*[c,d]

[a]*University of Sussex, Department of Psychology, Pevensey Building, Falmer, BN1 9QH, UK*
[b]*Behavioural Science Institute, Radboud University Nijmegen, P.O. Box 9104, 6500 HC Nijmegen, The Netherlands*
[c]*F.C. Donders Centre for Cognitive Neuroimaging, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands*
[d]*Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands*

ARTICLE INFO

Article history:
Accepted 23 February 2007
Available online 28 February 2007

Keywords:
N400
Discourse
World knowledge
Language
ERP

ABSTRACT

In an ERP experiment we investigated how the recruitment and integration of world knowledge information relate to the integration of information within a current discourse context. Participants were presented with short discourse contexts which were followed by a sentence that contained a critical word that was correct or incorrect based on general world knowledge and the supporting discourse context, or was more or less acceptable based on the combination of general world knowledge and the specific local discourse context. Relative to the critical word in the correct world knowledge sentences following a neutral discourse, all other critical words elicited an N400 effect that began at about 300 ms after word onset. However, the magnitude of the N400 effect varied in a way that suggests an interaction between world knowledge and discourse context. The results indicate that both world knowledge and discourse context have an effect on sentence interpretation, but neither overrides the other.

1. Introduction

Almost everyday we read or listen to new information. Listening to the news, speaking with a family member or co-worker, overhearing a conversation or watching a television show will likely all involve encountering some new information, such as an anecdote of something that happened that day. Deriving the meaning of each of these utterances involves, at the least, retrieving the meanings of the individual words and facts about the world and combining them in a way that is licensed by the grammar and the prior discourse context. However, we sometimes hear new facts that may or may not be congruent with what we already know, for example, when speaking with a friend (e.g., Mary says to Bob, "*John can't make it to the party*", when Bob thought John was coming to the party) or watching a television show (e.g., "*The pigs are flying low today, eh*", John Cleese, 1969 in Monty Python's Flying Circus). When this occurs, do we initially just incorporate the current discourse for understanding the sentence or do we also integrate our world knowledge, creating a conflict? The purpose of the current study is to investigate how the recruitment and integration of world knowledge information relate to the integration of information within a current discourse context. When the discourse information does not fit with our previous world knowledge, is the default world knowledge overwritten?

Models of language comprehension disagree on how and when discourse information and world knowledge are

* *Corresponding author.*
E-mail addresses: l.hald@sussex.ac.uk, lea.hald@gmail.com (L.A. Hald).

0006-8993/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2007.02.054

integrated during comprehension. There are several early theories of comprehension that assume that comprehension occurs in stages, and this assumption leads to two relevant implications: first, sentences (or constituents) are understood independently before being related to the prior discourse and secondly, because of this, discourse information is only available relatively later during processing (see Chomsky, 1975; Fodor et al., 1974; Forster, 1989; Katz, 1972; Searle, 1979; Sperber and Wilson, 1986). A less strong version of this idea suggests that discourse information does make contact with the ongoing processing after each word, however processes related to the discourse only occur after the individual word meaning, syntax and sentence-level semantics have been broadly established (Frazier, 1999; Fodor et al., 1996).

Alternatively, it has been pointed out that making a distinction between the integration of word meaning, discourse context and world knowledge is problematic (see Clark, 1996; Jackendoff, 2002). How can a single word or sentence not draw upon general background knowledge and discourse information? How do you draw the line between word meaning and world knowledge?

This alternative view is generally based on the idea of constraint satisfaction. The various types of information such as word meaning, syntax, discourse and world knowledge are all used simultaneously to reach the best interpretation of an utterance (Jackendoff, 2002; MacDonald et al., 1994; Marslen-Wilson and Tyler, 1980; Tanenhaus and Trueswell, 1995). However, for practical reasons, it is often the case that discourse information and world knowledge are actually not specified in these models, even though they are considered to play a role. One model that does specify the role of discourse and world knowledge is by Kintsch (1998).

In this model, comprehension involves building a mental representation of the incoming words at two levels, a local level and a global level. The local level processing includes tying sentence meanings together, such as relating anaphors to their referents. The global level processing includes establishing an overall meaning (i.e., the gist), specifically relating the sentences to the wider topic at hand. However, in order to reach a true understanding, a mental representation of the situation (i.e., situational model), which integrates world knowledge, is necessary. Although the first two levels are considered to occur automatically, this final step is assumed to require conscious effort.

There has been much recent research examining the integration of discourse information. For instance, in two different event related potential (ERP) studies, participants were presented with short discourses like the following: *As agreed upon, Jane was to wake her sister and her brother at five o'clock in the morning. But the sister had already washed herself, and the brother had even got dressed. Jane told the brother that he was exceptionally quick/slow*. In the final sentence, a critical word (underlined) was either congruent or incongruent with the discourse context. However, both of the critical words were equally plausible within the sentence context in isolation. An N400 effect (greater amplitude N400) was found for words that were anomalous within the discourse (but not anomalous within the sentence, i.e., *slow*). This N400 effect had the same time course, scalp distribution and overall morphology as N400 effects shown within single sentences, indicating that there is no fundamental difference between the integration of a word in its local (sentence level) and its global (discourse level) semantic context (Van Berkum et al., 1999, 2003). This result fits with a growing body of research suggesting that discourse information is integrated on-line as the sentence unfolds (see Garrod and Terras, 2000; Hess et al., 1995; Marslen-Wilson and Tyler, 1980; Spivey-Knowlton and Tanenhaus, 1998; St. George et al., 1994).

Another relevant finding is that the magnitude of the N400 effect can also inform us about the integration of discourse information. In an ERP study, Federmeier and Kutas (1999) visually presented participants with short discourses like the following: *They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of palms/pines/tulips*. Critically, each discourse either ended with a highly expected word (*palms*), a word that belonged to the same semantic category (e.g., trees) as the expected word (*pines*), or a word that came from a different semantic category (*tulips*). Although the authors found a discourse related N400 effect for both unexpected words, the critical result is that the magnitude of the N400 effect was smaller for the word belonging to the same semantic category compared to the word belonging to a different category. The authors interpret this difference as a reflection of the impact of the organization of long-term memory on language processing.

Results on the influence of world knowledge within sentences show a similar pattern of immediate integration as sentences unfold (Hald, 2003; Hald et al., 2004; Hagoort et al., 2004, Münte et al., 1998). Across two ERP experiments, participants either read or heard sentences like the following: *Amsterdam is a city that is very old/new/thin* and lively. The sentences were of three types: semantically well-formed and true, based on world knowledge (*old*), semantically well-formed, but not true (*new*), when considering the founding date of Amsterdam, or semantically not well-formed since the semantics of the noun *city* makes the semantic specification of the adjective thin not applicable. The results showed a clear and sizeable N400 effect, with equivalent onset and peak latencies for semantic and world knowledge violations (Hald, 2003; Hald et al., 2004; Hagoort et al., 2004).

The previous research on world knowledge does not support the model of Kintsch (1998), but is consistent with models that assume that both discourse and world knowledge information are immediately integrated during sentence processing. However, what is unclear from the current research is how discourse information and world knowledge information relate or interact.

The current experiment explores this issue. Following a short discourse context, one of two different sentence types is presented: a sentence that is correct based on knowledge of the world (critical words are underlined, World Knowledge Correct, WK+, e.g., *The city Venice has many canals and beautiful buildings*) or a sentence that violates world knowledge (World Knowledge Violation, WK-, e.g., *The city Venice has many roundabouts and beautiful buildings*). Importantly, these sentences are preceded by one of two different discourse contexts: a discourse containing information that supports the facts in the correct sentence (World Knowledge Compatible Discourse, Dwkc, e.g., *The city of Venice is surrounded by water. Many tourists like to go here and they love to take a gondola*

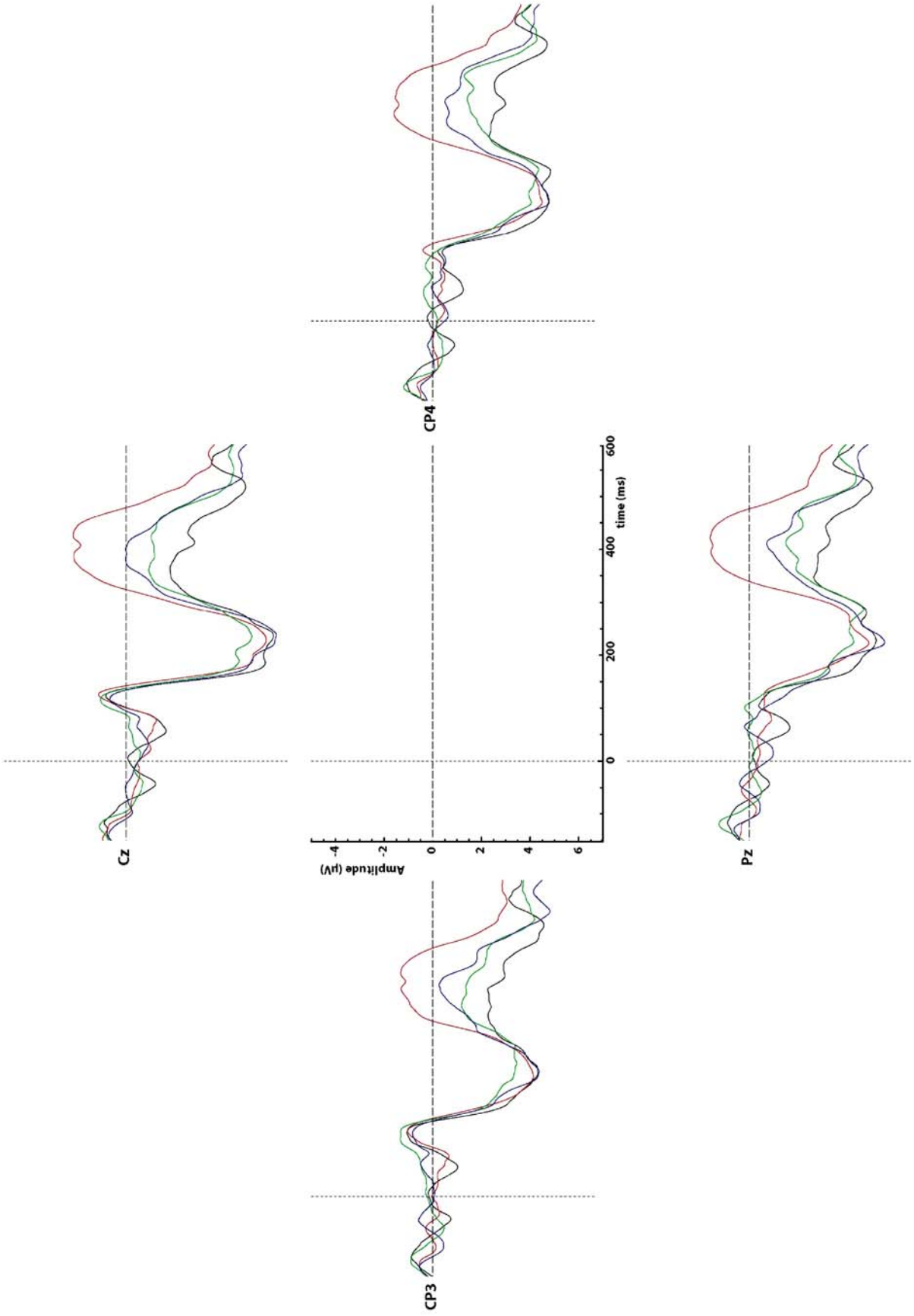

Fig. 1 – ERP waveforms at four centro-posterior electrode sites for the <Dwkc, WK+> condition (black line), the <Dwkc, WK–> condition (red line), the <Dldc, WK+> condition (green line) and the <Dldc, WK–> condition (blue line). The waveforms are time-locked (0 ms) to the critical word onset.

tour of the city. They are often impressed with the beauty of the city), or a discourse context that shifts the relevant focus, making the world knowledge violations more acceptable (Local Discourse Context, Dldc, e.g., *The large and increasing amount of cyclists in the inner city of Venice had to be regulated. The city council decided 10 years ago to replace traffic lights with other road layouts that ease traffic flow*). Comparing the ERP waveforms to each combination of discourse context and sentence type will let us look at the relationship between world knowledge and discourse information.

Previous ERP studies indicate that world knowledge and discourse information are involved in sentence processing very early, within 250–350 ms after word onset. This suggests that these forms of information probably interact at this early time point in order to arrive at the best interpretation of an utterance. If this is the case, each of these forms of information should be used in conjunction in a way that will be detectable by a modulation in the N400 effect. Specifically, when a discourse context shifts the relevant focus to support new information, a previous world knowledge violation should show a reduction in amplitude of the N400 (Dldc, WK-, e.g., summarized: *the large amount of cyclists in Venice caused the city council to replace traffic lights with other road layouts followed by The city Venice has many* roundabouts…). In this case, the discourse context will make the word *roundabouts* more acceptable and possibly easier to integrate into the ongoing message-level representation. Furthermore, when the same discourse context is followed by a sentence that is correct based on previous world knowledge (Dldc, WK+, e.g., *The city Venice has many* canals…), an increase in the N400 amplitude is expected. Here, the influence of the discourse context makes the correct world knowledge information less relevant or acceptable and more difficult to integrate. Finally, the largest amplitude N400 should be seen when neither the discourse context nor the background world knowledge supports the interpretation of the world knowledge violation (e.g., Dwkc, WK-, summarized: *The city of Venice is surrounded by water and many tourists go there. The city Venice has many* roundabouts…).

2. Results

Average waveforms were computed across all trials per condition for the 16 participants. The waveforms were normalized relative to a 150 ms pre-CW baseline. A latency window of 325–525 ms after CW onset was used to compute the mean amplitude of the N400.

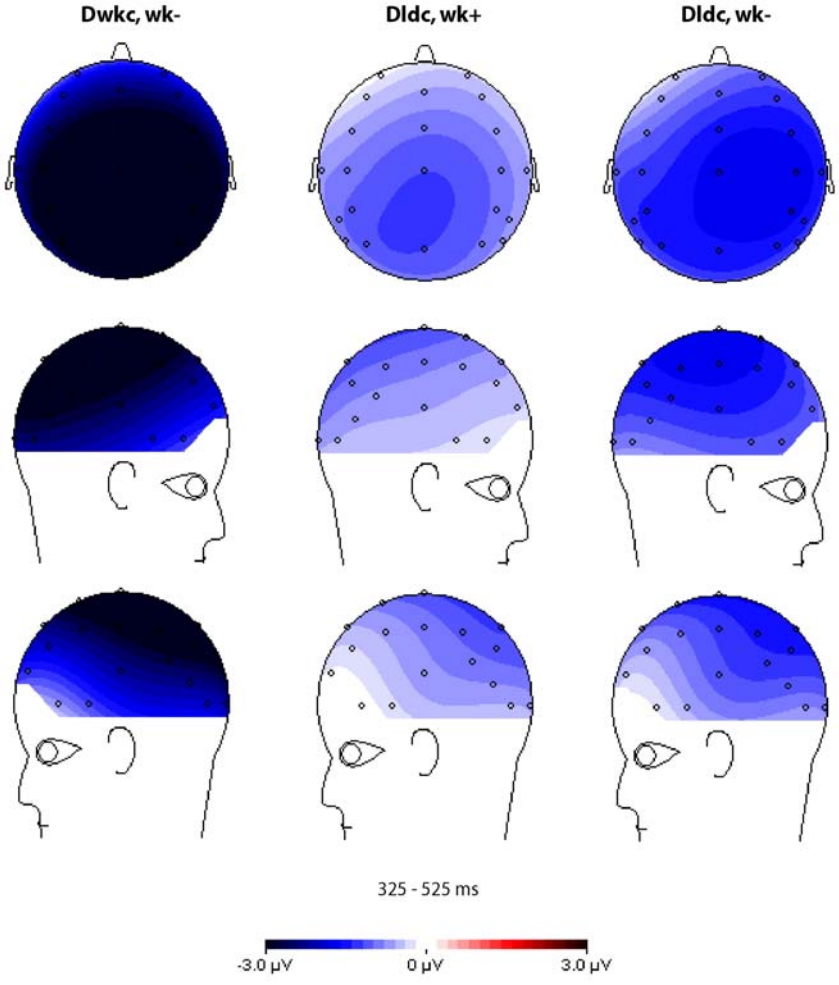

Fig. 2 – Scalp topography of the N400 for the <Dwkc, WK–>, <Dldc, WK+> and <Dldc, WK–> conditions for the time interval between 325 and 525 ms after critical word onset.

Results were first evaluated in an omnibus ANOVA, with WK (WK+, WK–), Discourse context (Dwkc, Dldc) and Electrode ($N=29$) as the within-subject factors. Planned comparisons were performed to look at the effect of WK within each context. Significant interactions were clarified by breaking them down into simple effects. F-tests that violated the sphericity assumption were adjusted by means of the Greenhouse–Geisser/Box's epsilon hat correction (Maxwell and Delaney, 1989).

Grand average ERP waveforms time-locked to the onset of the CW for all conditions are presented in Fig. 1 for four centro-posterior electrode sites.

From the grand average waveforms a clear N1–P2 complex is seen for the critical words, which is common for visually presented material. Beginning at about 300 ms the waveforms start to diverge and the conditions <Dwkc, WK–>, <Dldc, WK+> and <Dldc, WK–> elicit a negative deflection that peaks at about 400 ms. All three of these conditions show an N400 with the same onset and peak latency. However, the amplitude varies across conditions. The <Dwkc, WK–> has a larger amplitude N400 than do the <Dldc, WK+> and <Dldc, WK–> conditions. The distribution of the N400 shows the typical central–posterior maximum, as can be seen in Fig. 2.

The omnibus ANOVA on the mean amplitudes in the time window 325–525 ms latency range revealed a significant main effect of WK and a significant interaction between WK and Discourse context (see Table 1). There was no significant main effect of Discourse context. This analysis also revealed a significant three-way interaction between WK, Discourse context and Electrode ($F(28, 40.4)=4.12$; MSe = 8.07; $p=0.015$), due to the posterior distribution of the N400 effects.

The planned comparisons including all electrodes indicated a significant difference for the context which is compatible with our world knowledge (Dwkc) between the WK+ condition and the WK– condition. A smaller, but significant difference was also seen between the WK– condition in the local discourse context (Dldc) and the baseline. For the WK+ sentences, the effect of the discourse context was reversed, indicating a larger N400 in the local discourse context compared to the world knowledge compatible context. However, over all electrodes this effect was only marginally significant ($p<0.10$).

Since a three-way interaction was found between electrode site and the two experimental conditions, we conducted an omnibus ANOVA on the mean amplitudes in the time window 325–525 ms for just the posterior electrodes, where the effects were the greatest.

As shown in Table 1, at posterior electrode sites, we found the same pattern of results as for the main effects. The main effect of WK was significant, as well as the interaction of WK and Discourse context. Furthermore, again a significant difference was found in the Dwkc context between the WK + condition and the WK– condition, with a sizeable increase in amplitude for the N400 in the latter condition (difference: −2.89 μV). Similarly, a significant difference was again seen between the WK– condition in the local discourse context (Dldc) and the baseline condition (difference: −1.25 μV). Moreover, over the posterior electrodes the WK+ sentences showed a significantly larger N400 in the local discourse context compared to the world knowledge compatible context (difference: −0.83 μV). The contrast <Dldc, WK+> versus <Dldc, WK–> was not significant ($p=0.416$). No other significant differences were seen in the waveforms at the critical word, other than the N400 effects. Furthermore, no differences were seen in the word preceding the critical word.

Table 1 – Main effects and planned comparisons for the N400 effects in the 325 to 525 ms latency range

Source	F	MSe	p
Main effects with all 29 electrodes			
Discourse context	1.3	39.39	0.27
WK	21.93	22.00	<0.001***
Discourse × WK	21.49	48.87	<0.001***
Planned comparisons with all 29 electrodes			
Dwkc, WK– [a]	39.28	75.27	<0.001***
Dldc, WK– [a]	6.61	96.35	0.021
Dldc, WK+ [a]	3.09	70.92	0.099
Main effects, posterior electrodes			
Discourse context	2.03	15.02	0.175
WK	22.06	13.55	<0.001***
Discourse × WK	17.50	30.30	<0.001***
Planned comparisons, posterior electrodes			
Dwkc, WK– [a]	39.97	40.65	<0.001***
Dldc, WK– [a]	6.00	51.04	0.027*
Dldc, WK+ [a]	4.74	29.21	0.046*

Mean amplitudes and effect sizes (d) at posterior electrodes ($N=12$)

Source	Mean	d
Dwkc, WK+ (baseline)	3.01 μV	
Dwkc, WK–	0.12 μV	−2.89 μV
Dldc, WK–	1.76 μV	−1.25 μV
Dldc, WK+	2.18 μV	−0.83 μV

[a] Comparisons are made relative to the baseline condition Dwkc, WK+. Posterior electrode analysis included electrodes LTP, LP, CP3, P3, P07, RTP, RP, CP4, P4, P08, Pz and Oz.

3. Discussion

We conducted an ERP experiment to investigate how world knowledge information relates to the integration of discourse context. In the experiment, participants were presented with short discourse contexts that either corresponded to general world knowledge (e.g., World-Knowledge Compatible Discourse: *The city of Venice is surrounded by water. Many tourists like to go here and they love to take a gondola tour of the city. They are often impressed with the beauty of the city*), or did not correspond to general world knowledge and instead changed the focus to suggest new or different information (e.g., Local Discourse: *The large and increasing amount of cyclists in the inner city of Venice had to be regulated. The city council decided 10 years ago to replace traffic lights with other road layouts that ease traffic flow*). These short contexts were followed by sentences that were either correct based on general world knowledge (World Knowledge Correct, WK+: *The city Venice has very many canals and beautiful buildings*) or incorrect based on general world knowledge but more

acceptable relative to the local discourse (World Knowledge Violation, WK–: *The city Venice has very many roundabouts and beautiful buildings*). Relative to the baseline condition in which both the spoken discourse context and the information in the visually presented sentence were consistent with world knowledge in long-term memory, the other three conditions all elicited an N400 (i.e., a negative shift in the waveforms that began around 300 ms and peaked around 400 ms). In addition, the magnitude of the N400 effect reflected an interaction between world knowledge information and discourse context. This interaction was substantiated statistically. The largest amplitude N400 was seen when the sentence contained a world knowledge violation that was both incongruent with the current discourse context and world knowledge in long-term memory. However, when the discourse context provided a more acceptable interpretation for the world knowledge violation, the N400 amplitude was reduced. Moreover, when current discourse context instantiated a focus shift from the general world knowledge, the amplitude of the sentence that was compatible with our world knowledge increased relative to the same sentence in a supporting discourse context. The results found when the discourse context supported real world knowledge information corresponded with the cloze probabilities found for the congruent versus incongruent words in this context. However, this was not the case when the discourse context provided a more acceptable interpretation of the world knowledge violation. Here we see no difference between the two sentence conditions, even when there was a difference in the cloze ratings (28.12% versus 58.37%).

These results indicate that there is an interaction between the world knowledge information in long-term memory and the local discourse context. Local discourse can provide a context for new information to be more acceptable. However, the local discourse context does not completely override the world knowledge information in long-term memory. Furthermore, world knowledge information in long-term memory does not completely override the local discourse. Even "correct" world knowledge information can be less acceptable if the local discourse context does not support that world knowledge.

Both world knowledge and discourse information conspire during the ongoing sentence processing as it unfolds in real time. That is, very rapidly and within the same time window both local discourse and general world knowledge information are taken into consideration during the construction of a sentence interpretation. This interaction allows for the comprehension system to pursue an interpretation that is more inline with the discourse context, even when that information does not fit facts about the world (see also Nieuwland and Van Berkum, 2006 for a similar finding with cartoonlike stories). Our finding supports the idea that various types of information such as word meaning, syntax, discourse and world knowledge are all recruited simultaneously when interpreting an utterance (see Clark, 1996; Jackendoff, 2002; MacDonald et al., 1994; Marslen-Wilson and Tyler, 1980; Tanenhaus and Trueswell, 1995). The results suggest that this process includes the interaction of at least world knowledge and discourse. As discussed in the Introduction section, it is often the case that discourse information and world knowledge are actually not specified in processing models, even when they are considered to play a role. An exception to that is Kintsch (1998), who does specify the role of discourse and world knowledge during processing. Discourse is considered to be included in the global level of processing in this model, which is one of two levels of processing considered to be fairly automatic. However, this model further assumes that world knowledge necessary to create a mental representation of the situation, which is needed in order to reach a true understanding of an utterance, requires conscious effort. Although the early effects of discourse information we found generally correspond to this model, the finding of an early influence of world knowledge and the interaction of discourse context and world knowledge do not correspond with the idea that world knowledge is only incorporated in a final conscious step.

In summary, our findings lead to three related conclusions about how the recruitment and integration of world knowledge information and the integration of discourse information conspire. First, world knowledge information from long-term memory and local discourse context interact during on-line sentence comprehension. Secondly, local discourse can provide a context for differing or new world knowledge information to be more acceptable and integrated easier. Finally, although local discourse can have an effect on world knowledge integration, it does not override world knowledge information in long-term memory, nor can the information in long-term memory completely override the local discourse context.

4. Experimental procedures

4.1. *Participants*

Twenty-seven native speakers of Dutch from the F.C. Donders Centre for Cognitive Neuroimaging subject pool participated in the experiment, sixteen of which were included in the final analysis (12 females; aged 19–42, mean = 23.8). They were paid a small fee for their participation. All participants had normal or corrected-to-normal vision, normal hearing and all were right handed. None of the participants had any neurological impairment nor had any of the participants participated in the pretests (see below). Eleven of the participants that participated were excluded from the final analysis due to excess eye movement, excessive noise from muscle tension or technical problems with recording (see EEG recording and analysis below).

4.2. *Stimulus material*

The experimental materials consisted of 160 pairs of critical sentences in Dutch. The pairs of critical sentences were identical with the exception of one word (critical word, CW, in capital letters in Table 2). In addition, all pairs contained a topic word or words that were proper names of people, places, events or things that were used to examine people's knowledge of those things. Each pair consisted of a sentence that was completely semantically coherent and correct based on knowledge of the world (world knowledge correct sentence, WK+), and a second sentence that was semantically coherent but false based on world knowledge (world

Table 2 – Example of the materials

Discourse contexts	
Dwkc	*De stad Venetie is omgeven door water. Veel toeristen komen er graag om met een gondel rond te varen in de stad. Ze zijn vaak onder de indruk van de schoonheid van de stad.* (The city of Venice is surrounded by water. Many tourists like to go here and they love to take a gondola tour of the city. They are often impressed with the beauty of the city.)
Dldc	*De grote hoeveelheid fietsers in de binnenring van Venetië moet in goede banen geleid worden. Daarom heeft de overheid 10 jaar geleden besloten om de stoplichten te vervangen door betere wegconstructies die de doorstroom bevorderen.* (The large and increasing amount of cyclists in the inner city of Venice had to be regulated. The city council decided 10 years ago to replace traffic lights with other road layouts that ease traffic flow.)
Critical sentences	
Wk+	*De stad Venetië heeft heel veel grachten en mooie gebouwen.* (The city Venice has very many canals and beautiful buildings.)
Wk–	*De stad Venetië heeft heel veel rotondes en mooie gebouwen.* (The city Venice has very many roundabouts and beautiful buildings.)

The contexts were presented auditorily. The critical sentences were presented visually. The Critical Words are underlined.

knowledge violation, WK–). In addition, for each pair of sentences, two discourse contexts were created. All of the discourse contexts consisted of either 2 or 3 sentences (held consistent within pairs of discourse contexts, 120 contexts were 2 sentences, 40 contexts were 3 sentences). Specifically, one of the discourse contexts was compatible with world knowledge (Dwkc). The compatible discourse gave supporting context to the information described in the WK+ sentence, supporting our actual world knowledge. In contrast, the other discourse changed the relevant focus so that the WK– sentence became more acceptable. We refer to this discourse context as the local discourse context (Dldc). By combining the sentences and the discourses orthogonally, a fully factorial design was realized, with the <Dwkc, WK+> condition as the baseline.

In order to avoid lexical priming from the discourses to the CWs, none of the discourse contexts contained the CW itself, a synonym of the CW or a semantically or associatively related lexical item to the CW. Words in the critical sentences that were not the CW were either mentioned in both discourse contexts or not mentioned in either discourse context. An example of the materials is found in Table 2.

The discourse contexts were recorded by a female speaker at a normal speaking rate with normal intonation. The average length of the discourses was 10.65 s with a standard deviation of 2.64 s. The critical sentences were presented visually (see Procedure for details).

For the critical sentences, the CW was never in the sentence-final position and always appeared before the 11th word in the critical sentence. Sentences varied in length from 5 to 14 words, with the average sentence length being 9 words long (SD=1.9). The critical words were matched across conditions on the following criteria: (i) word frequency did not differ significantly (WK+: 1.5; WK–: 1.4), based on log lemma frequencies (CELEX, Baayen et al., 1993); (ii) word class (equated within each pair). Singular nouns were of the same gender. The critical words were not ambiguous. None of the critical words were over 12 letters in length.

The 160 experimental items were pseudorandomized across 4 different versions of the experiment so that each participant heard/read only one of the discourse/critical sentence pairs. Each experimental version contained 40 exemplars of each of the 4 conditions. The ERP experiment was split into 4 blocks lasting approximately 13 min each. In addition, there was a practice block of 20 items and 8 extra starter items, which were similar in nature to the experimental items. These starter items were used two at a time at the start of each experimental block to minimize loss of data due to artifacts after beginning a new block. Following each block there was a short break. No other fillers were used.

Two pretests (described below) were performed on these materials. The purpose of these pretests was twofold: first, they would assess the cloze probability of the CW. Secondly, the first pretest was additionally used to ensure that the knowledge necessary for the correct sentence was knowledge that the majority of the participants in the subject pool would share.

4.3. *Pretests*

The first pretest was conducted to establish that people were knowledgeable about the topics in our materials and thus would realize the world knowledge violations. Furthermore, this pretest established the cloze probability of the CWs. The cloze probability test requires participants to complete sentence fragments, and the "cloze probability" refers to the proportion of participants who completed a particular sentence fragment with a particular word (Taylor, 1953).

For each sentence pair, one sentence that stopped just prior to the critical word was given and the participants were required to complete the sentence with one word (e.g., *The city Venice has many* _____). The sentences were pseudorandomized to ensure that sentences with similar topics were not presented in immediate succession.

Sixteen native speakers of Dutch from the F.C. Donders Centre for Cognitive Neuroimaging subject pool completed this cloze test (7 females; mean age=32). The mean cloze probability for the correct CWs was 51.29% (SD=32.38; range 12.5%–100%). In addition, no participant completed the sentences with the critical words that were used to create the world knowledge violations (or similarly incorrect words). We took this to indicate that participants were knowledgeable about the topics used to create these violations. For example, if any participant completed the sentence *The city Venice has many* _____ with the word *skyscrapers* or any other such word that would complete the sentence incorrectly, then we would assume that they do not have the world knowledge necessary about this topic. Additionally, to ensure that participants were knowledgeable about the topics in the sentences, all sentences needed to be completed with the CW by at least 12% of the participants. An item was replaced or changed from the item set if it either was not completed with the CW by 12% of the participants or if any participant completed the sentence incorrectly based on world knowledge.

A second pretest was performed on these critical sentences when they were preceded by either the Dwkc or Dldc contexts. The purpose of this second pretest was to determine whether people would change their preference after considering the relevant discourse context (i.e., would the cloze probability for the CWs change?).

Like in the first pretest, for each sentence pair, one sentence that stopped just prior to the critical word was given and the participants were required to complete the sentence with one word (e.g., *The city Venice has many* _____). However, in this case the participants would also either read the world knowledge compatible context or the local discourse context prior to the truncated critical sentence. Participants were instructed to complete the truncated sentence with one word that would be appropriate to the story given. The items were pseudorandomized to ensure that similar topics were not presented consecutively.

Sixteen native speakers of Dutch from the F.C. Donders Centre for Cognitive Neuroimaging subject pool who did not participate in the previous pretest completed this second cloze test (6 female; mean age=27). The results of this cloze procedure were divided into 3 categories: i. responses that were the same or similar semantically to the WK+ CW; ii. responses that were the same or similar semantically to the WK– CW; iii. responses that could not be classified in either of the first two categories. Table 3 illustrates the cloze probabilities for each discourse context.

We established two criteria for accepting items for the final ERP experiment. First, after the presentation of a Dwkc, the cloze probability for the WK+ CW needed to reach at least 50%, which was found in the data (mean=85.44%, SD=17.39, range 50%–100%). Furthermore, after the presentation of a Dldc, the cloze probability for the WK– CW needed to reach at least 25%, indicating that the local discourse could indeed shift people's preferences in such a way that they would be more likely to accept the world knowledge violation. In 91 of the 160 critical sentences, the WK– CW was given as the next word in the sentence when presented in the Dldc condition. This was also indicated in the data (mean=58.37%, SD=21.58, range 25%–100%).

The results of these two cloze tests indicate that participants did have the relevant knowledge about the topics in the items and would be able to realize a world knowledge violation. Furthermore, the results of the second cloze test, when discourse contexts preceded the truncated critical sentences, indicates that the Dwkc provided a good context for the correct sentence and the Dldc provided a context that made the world knowledge violation more acceptable based on the new information in the current discourse.

Table 3 – Results of the second cloze procedure containing the discourse contexts followed by a truncated critical sentence across all three categories

Discourse presented	Cloze probability by category		
	Category 1: WK+ CW	Category 2: WK– CW	Category 3: unrelated to either CW
Dwkc	85.44% SD=17.39	0.78% SD=4.13	13.78% SD=16.38
Dldc	28.12% SD=20.80	58.37% SD=21.58	13.52% SD 18.44

4.4. *Procedure*

Participants were tested individually in sound-attenuating, electrically shielded booth. The booth was dimly illuminated (Fiber optic lights DMX 512 at 40%). Participants were seated in a comfortable reclining chair and were told that the experiment was investigating how people integrate listening to and reading a story and that some of the sentences would be more difficult or stranger than other sentences. Therefore, they were unaware of the actual purpose of the study. They were told that they would hear a number of short stories that would each be followed by a written sentence that would be presented word-by-word in the middle of the computer screen. Participants were asked to attentively listen to and read the sentences and try to understand them as well as possible. They were also asked to try not to move or blink during the presentation of the sentence on the computer screen (the critical sentence). No other task demands were imposed.

Every trial began with a discourse context presented over speakers, during which time an asterisk was displayed in the middle of the computer screen for participants to focus on. At 1000 ms after the offset of the spoken discourse, the critical sentence was presented word-by-word in white lowercase Arial letters (18-point font size). The first word of each sentence was capitalized and the final word of each sentence was presented with a period. The letters were displayed against a dark background in the center of a VGA 60 Hz computer screen. Viewing distance was approximately 110 cm, and the largest word subtended a visual angle of about 3.2° horizontally and 0.5° vertically. Each word was presented for 300 ms followed by a blank screen for 300 ms that was followed by the next word. After the final word there was a blank screen for 2 s that was followed by an asterisk displayed in the middle of the screen for 1 s, indicating to the participants that they could blink and move their eyes. There was a 300 ms delay between the asterisk disappearing and the start of the next trial. Sentences were presented using the Nijmegen Experimental Set-Up software (NESU; http://www.mpi.nl/world/tg/experiments/nesu.html).

4.5. *EEG recording and analysis*

The EEG was recorded from 29 Ag/AgCl-sintered electrodes mounted in a cap, each referred to the left mastoid. Five electrodes were placed according to the 10%-standard system of the American Electroencephalographic Society over midline sites at Fz, FCz, Cz, Pz and Oz locations, along with 9 lateral pairs of electrodes over standard sites on frontal (AF3, AF4, F3, F4, F7 and F8), fronto-central (FC3 and FC4), fronto-temporal (FT7 and FT8), central (C3 and C4), centro-parietal (CP3 and CP4), parietal (P3 and P4) and occipital (PO7 and PO8) positions. Three additional pairs were placed laterally over symmetrical positions: (a) a temporal pair (LT and RT) placed laterally to Cz, at 33% of the interaural distance, (b) a temporo-parietal pair (LTP and RTP) placed 30% of the interaural distance lateral and 13% of the nasion–inion distance posterior to Cz, and (c) a parietal pair midway between LTP/RTP and PO7/PO8 (LP and

RP). Vertical eye movements were monitored via a supra- to sub-orbital bipolar montage. A right to left canthal bipolar montage was used to monitor for horizontal eye movements. Activity over the right mastoid bone was recorded on an additional channel to determine if there were differential contributions of the experimental variables to the presumably neutral mastoid site. No such differential effects were observed.

The EEG and EOG recordings were recorded with two 32-channel BrainAmp DC amplifiers from Brain Products, using a hi-cut of 70 Hz (no notch filter) and a time constant of 10 s (0.016 Hz). Electrode impedances were kept below 3 kΩ for the EEG recording and below 5 kΩ for the EOG recording. The EEG and EOG signals were digitized on-line with a sampling frequency of 500 Hz. The signals were amplified and stored on a dedicated data-processing computer system for further off-line analysis. The software package Vision Analyzer of Brain Vision was used to analyze the waveforms.

The EEG data were screened for eye movements, electrode drifting, amplifier blocking and EMG artifacts in a critical window ranging from 150 ms before to 1200 ms after the onset of the critical word. Trials containing such artifacts were rejected (10.75% overall). Six participants were excluded from the analysis because more than 20% of the trials were rejected due to excessive eye blinks or muscle tension. Five remaining participants were excluded due to experimenter error or technical problems with recording. Thus, 16 participants remained for subsequent analysis.

Acknowledgments

We are grateful to Miriam Kos for her assistance in running the subjects. For an early part of this project, Lea Hald was funded by a U.S. Fulbright fellowship.

REFERENCES

Baayen, R.H., Piepenbrock, R., van Rijn, H., 1993. The CELEX lexical database (CD-ROM) [Computer Software]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Chomsky, N., 1975. Reflections on Language. Pantheon, New York, NY.

Clark, H.H., 1996. Using Language. Cambridge Univ. Press, Cambridge.

Federmeier, K.D., Kutas, M., 1999. A rose by any other name: long-term memory structure and sentence processing. J. Mem. Lang. 41, 469–495.

Fodor, J.A., Bever, T.G., Garrett, M.F., 1974. The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. McGraw-Hill Book Company, New York, NY.

Fodor, J.D., Ni, W., Crain, S., Shankweiler, D., 1996. Tasks and timing in the perception of linguistic anomaly. J. Psycholinguist. Res. 25, 25–57.

Forster, K.I., 1989. Levels of processing and the structure of the language processor. In: Cooper, W.E., Walker, C.T. (Eds.), Sentence Processing. Erlbaum, Hillsdale, NJ.

Frazier, L., 1999. On Sentence Interpretation. Kluwer, Dordrecht, NL.

Garrod, S., Terras, M., 2000. The contribution of lexical and situational knowledge to resolving discourse roles: bonding and resolution. J. Mem. Lang. 42, 526–544.

Hagoort, P., Hald, L., Bastiaansen, M., Peterson, K.M., 2004. Integration of word meaning and world knowledge in language comprehension. Science 304 (5669), 438–441.

Hald, L., 2003. The integration of semantic versus world knowledge during on-line sentence comprehension. (Doctoral dissertation, University of Arizona, 2002). Dissertations Abstracts International, DAI-B 63/12, 6114.

Hald, L., Hagoort, P., Van Galen, M., 2004. Semantic processing versus world knowledge processing during on-line language comprehension: an auditory ERP study. Cogn. Neurosci. Soc. Abstr. 11.

Hess, D.J., Foss, D.J., Carroll, P., 1995. Effects of global and local context on lexical processing during language comprehension. J. Exp. Psychol. Gen. 124 (1), 62–82.

Jackendoff, R., 2002. Foundations of Language. Oxford Univ. Press, New York, NY.

Katz, J., 1972. Semantic Theory. Harper and Row, New York, NY.

Kintsch, W., 1998. Comprehension: A Paradigm for Cognition. Cambridge Univ. Press, Cambridge.

MacDonald, M.C., Pearlmutter, N.J., Seidenburg, M.S., 1994. Lexical nature of syntactic ambiguity resolution. Psychol. Rev. 101 (4), 676–703.

Marslen-Wilson, W.D., Tyler, L.K., 1980. The temporal structure of spoken language understanding. Cognition 8, 1–71.

Maxwell, S.E., Delaney, H.D., 1989. Designing and Analyzing Data. Brooks/Cole, Pacific Grove, CA, pp. 475–479.

Münte, T.F., Schiltz, K., Kutas, M., 1998. When temporal terms belie conceptual order. Nature 395, 71–73.

Nieuwland, M.S., Van Berkum, J.J.A., 2006. When peanuts fall in love: N400 evidence for the power of discourse. J. Cogn. Neurosci. 18 (7), 1098–1111.

Searle, J., 1979. Literal meaning. In: Searle, J. (Ed.), Expression and Meaning. Cambridge Univ. Press, Cambridge.

Sperber, D., Wilson, D., 1986. Relevance. Harvard Univ. Press, Cambridge, MA.

Spivey-Knowlton, M.J., Tanenhaus, M.K., 1998. Syntactic ambiguity resolution in discourse: modeling the effects of referential context and lexical frequency. J. Exper. Psychol., Learn., Mem. Cogn. 24, 1521–1543.

St. George, M., Mannes, S., Hoffman, J.E., 1994. Global semantic expectancy and language comprehension. J. Cogn. Neurosci. 6 (1), 70–83.

Tanenhaus, M.K., Trueswell, C., 1995. Sentence comprehension. In: Miller, J.L., Eimas, P.D. (Eds.), Speech, Language, and Communication. Academic Press, San Diego, CA.

Taylor, W.L., 1953. "Cloze procedure": a new tool for measuring readability. Journal. Q. 30, 415–433.

Van Berkum, J.J.A., Hagoort, P., Brown, C., 1999. Semantic integration in sentences and discourse: evidence from the N400. J. Cogn. Neurosci. 11 (6), 657–671.

Van Berkum, J.J.A., Zwitserlood, P., Hagoort, P., Brown, C., 2003. When and how do listeners relate a sentences to the wider discourse? Evidence from the N400 effect. Cogn. Brain Res. 17, 701–718.