

Discourse expectations are sensitive to the Question under Discussion: Evidence from
ERPs

Francesca Delogu, Torsten Jachmann, Maria Staudte
Saarland University

Francesco Vespignani
University of Trento

Nicola Molinaro
BCBL, Basque Center on Cognition, Brain, and Language

Author Note

Correspondence concerning this article should be addressed to Francesca Delogu,
Department of Language Science & Technology, Building C7.1, Saarland University,
66123 Saarbruecken, Germany. Email: delogu@coli.uni-saarland.de

Abstract

Questions under Discussion (QUDs) have been suggested to influence the integration of individual utterances into a discourse-level representation. Previous work has shown that processing ungrammatical ellipses is facilitated when the elided material addresses an implicit QUD raised through a non-actuality implicature (NAIs; Grant et al., 2013). It is not clear, however, if QUDs influence discourse coherence during comprehension of fully acceptable discourse. We present two ERP studies examining the effects of QUDs introduced by NAIs using two-sentence discourses. Experiment 1 showed that processing definite NPs with inaccessible antecedents is facilitated when their content is relevant to the QUD. Using acceptable discourses, Experiment 2 showed that definite NPs failing to address a QUD elicit increased processing cost. Overall, our results indicate that QUDs raise the expectation that the following discourse will address them, providing unambiguous evidence that their influence is not limited to the processing of ungrammatical input.

Keywords: Question Under Discussion; Implicatures; Discourse coherence; Discourse comprehension; ERP

Discourse expectations are sensitive to the Question under Discussion: Evidence from ERPs

Introduction

Discourse comprehension requires readers to build a coherent mental representation of what is being communicated by relating any new sentence to previously processed information. A considerable amount of work in (psycho-)linguistics has been devoted to determine which factors contribute to discourse coherence and to characterize the different types of coherence relations that connect clauses and sentences in discourse (e.g., Hobbs, 1985; Kehler, 2002; Kintsch & van Dijk, 1978; Sanders & Noordman, 2000; van den Broek, 1990). A different - perhaps more fundamental - account of discourse coherence is based on the notion of Question under Discussion (Roberts, 1996, 2004, henceforth QUD). In this approach, discourses are organized by question/answer relations, with questions being often implicit and inferable on the basis of different cues (see also Carlson, 1983; Kuppevelt, 1995). For example, a sentence like “John was thirsty” may raise a potentially large number of implicit QUDs (e.g., Why? How did he get thirsty? What happened next? Did he get something to drink?), such that a subsequent sentence that addresses one of these questions (e.g., “He opened the fridge and grabbed a beer”) is perceived as a coherent continuation (e.g., Clifton & Frazier, 2012; Kuppevelt, 1995).¹

In this example, the QUD implicitly addressed by the continuation sentence is accommodated only after the content of the sentence has been processed. A growing body of evidence, however, suggests that comprehenders can use contextual cues to form expectations about specific QUDs that the following discourse should address, and that such expectations can guide the processing of subsequent discourse material (Clifton & Frazier, 2012, 2017; Grant, Clifton, & Frazier, 2012; Kehler & Rohde, 2017).

Kehler and Rohde (2017), for example, presented participants with sentences

¹ While there have been several attempts to integrate the QUD approach with theories based on coherence relations, the work is still in progress (Benz & Jasinskaja, 2017). We are therefore agnostic as to whether QUDs subsume coherence relations or are rather complementary to them.

containing implicit causality verbs (as in “Mary *scolded* John”) vs. control verbs (e.g., “Mary *saw* John”) and found that implicit causality contexts are more likely to elicit continuations answering a *Why?* question, whereas control verbs are biased toward answering other types of questions (e.g., *What happened next?*). While Kehler and Rohde (2017)’s findings suggest that comprehenders can generate expectations about a wide range of QUDs at least offline, Clifton and Frazier (2012) argued that the number of QUDs considered during online processing is more limited. Their view is based on a widely accepted semantic analysis of questions as denoting a set of possible (Hamblin, 1973) or true (Karttunen, 1977) answers. As any other question, an explicitly or implicitly introduced QUD is just the set of alternatives that potentially answer the question. Whenever the following discourse selects from among these alternatives, discourse coherence is increased and comprehension facilitated. According to Clifton and Frazier, there is a limited inventory of syntactic and pragmatic devices that set up a set of alternatives introducing a QUD. These include overt questions, contrastive focus, indefinites, disjunctions, as well as a novel type of conversational implicature termed non-actuality implicature (Frazier, 2008; Grant et al., 2012, henceforth NAI), which is the focus of the present paper.

According to the NAI hypothesis (Frazier, 2008), modal operators such as *should have* and *could have* as well as so called *intensional* verbs such as *want* and *need* pragmatically imply that the state of affairs described does not hold in the actual world. For example, a sentence like “*John wants (to drink) a beer*” implicates that a certain goal state (e.g., {John drinks/has a beer}) is not actual. This implicature is argued to introduce a QUD denoted by two alternatives ({*John gets a beer; John does not get a beer*}), thereby making any discourse continuation selecting one of the alternatives easier to process (*cf.* Grant et al., 2012). In line with this hypothesis, Grant and colleagues found that sentences containing elliptic material that mismatches syntactically with its antecedent are rated as more acceptable and read faster when the elliptic material comments on the QUD introduced by a NAI (e.g., “*This information needed to be released, but Gorbachev didn’t*”) compared to when the same material

follows a context not raising the QUD (e.g., “*This information was released, but Gorbachev didn’t*”). Similar results were found by Clifton and Frazier (2012) in a series of eye-tracking experiments demonstrating reading time cost for sentential material that fails to address the QUD introduced in the initial clause.

These findings suggest that the online processing of expressions that answer - or comment on - a QUD introduced by a NAI is facilitated, especially when the input is ungrammatical or unacceptable, as it is arguably the case for mismatch ellipsis (see also Clifton & Frazier, 2017). It is not clear, however, to what extent such facilitations are limited to the processing of linguistic constructions showing reduced acceptability (if not ungrammaticality) or are obtained as part of a more general mechanism guiding online discourse comprehension. The main goal of the present study was to investigate the influence of QUDs on comprehenders expectations about upcoming material in fully acceptable discourses. Following Clifton and Frazier (2017), we argue that QUDs exert their influence on the ongoing interpretation of the discourse and not just on a later stage in which the already computed meaning of a sentence is integrated with prior discourse (cf. Clifton & Frazier, 2017, p. 111). We therefore hypothesize that once a specific question is accepted as the QUD, comprehenders expect the incoming material to address it and interpret the current input in the context of such expectation. Thus, comprehenders will not wait until the end of the current sentence to relate it to the QUD, but they will incrementally interpret each chunk in relation to it.

We addressed these hypotheses in two event-related brain potential (ERP) experiments using two-sentence discourses in which we manipulated whether or not the context sentence introduced a QUD via a NAI, and whether or not the second sentence started with information relevant to the QUD. The first experiment still considered problematic input and investigated whether processing anaphoric definite NPs with inaccessible antecedents is facilitated when the content of the anaphor is relevant to the QUD. Experiment 2 investigated the effects of QUDs on the processing of definite NPs that could or could not be interpreted as initiating a comment on the QUD in fully acceptable discourses, as determined through a norming study.

Given the relative lack of prior ERP research on this topic, it is not clear which ERP components might be modulated by whether or not a NP addresses an implicit QUD. Two ERP components that have received the most attention in the investigation of discourse comprehension are the N400 and the P600. The N400 is a negative-going wave peaking approximately 400 ms post stimulus onset, more pronounced over centro-parietal sites. The amplitude of the N400 is sensitive to a variety of factors, including the degree to which a word is expected (and its meaning pre-activated) given the preceding sentential or discourse context (e.g., Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Lau, Phillips, & Poeppel, 2008, see Kutas & Federmeier, 2011 for a review). Particularly relevant to the present study are the findings of Burkhardt (2006), who observed a modulation of the N400 as a function of the degree of relatedness between a definite NP (e.g., *The conductor*) and prior information. Entities that form an identity relation with a previously introduced referent (e.g., *Tobias visited a conductor in Berlin*) elicit a reduced N400 amplitude compared to entities that can be linked though an inferential bridging relation (e.g., *Tobias visited a concert in Berlin*), which, in turn, elicit a reduced N400 amplitude compared to unrelated entities (e.g., *Tobias talked to Nina*). Interestingly, unrelated and bridged NPs also produced a P600 effect (see also Burkhardt, 2007). The P600, a positive deflection starting at around 500 ms post-stimulus onset and lasting several hundred milliseconds, has traditionally been associated with syntactic reanalysis or repair (e.g., Hagoort, Brown, & Groothusen, 1993). Burkhardt's findings, however, add to a large number of others (see Brouwer, Fitz, & Hoeks, 2012, for a review) pointing to an interpretation of the P600 as a general index of semantic integration cost, reflecting processes of revision or updating of the discourse model (Brouwer, Crocker, Venhizen, & Hoeks, 2017; Brouwer et al., 2012; Burkhardt, 2006, 2007; Spotorno, Cheylus, Henst, & Noveck, 2013). Based on these findings, a straightforward prediction is that lexical material that is interpreted as commenting on an implicit QUD should be easier to retrieve (N400) as well as to integrate (P600) into the unfolding discourse representation compared to material that cannot be easily interpreted as commenting on the QUD.

Experiment 1

A crucial part of forming a coherent discourse representation is resolving anaphoric expressions such as pronouns or definite NPs. To understand a short discourse like “*John drank a beer. [The beer/it] was warm*”, the reader has to establish a co-reference link between the anaphoric expression (*the beer, it*) and the indefinite NP antecedent (*a beer*). Indefinite NPs in the scope of intensional verbs (as in “*John wanted a beer*”), however, are inaccessible as antecedents for definite anaphora (i.e. for anaphors located in a factual clause), since the indefinites receive a special non-referential reading under which the existence of a referent is no longer presupposed (e.g., Moltmann, 1997; Van Geenhoven & McNally, 2005). Processing a pronoun or a definite NP in such contexts is predicted to engender a cost reflecting comprehenders’ attempt to restore coherence. The few studies examining reference processes in such contexts, however, have produced mixed results.

Using ERPs, Dwivedi, Phillips, Lague-Beauvais, and Baum (2006) observed a P600-like effect at the verb following the pronoun in a sentence like “*It ends quite abruptly*” when the context sentence introduced a hypothetical object (“*John is considering writing a novel*”) as opposed to when it did not (“*John is writing a novel*”). This effect was interpreted as reflecting a more costly resolution stage in which the context sentence is structurally re-analyzed to obtain a specific reading of the indefinite NP (e.g., *John is considering writing a [specific] novel [about his past failed romances]*).²

This penalty, however, was not replicated in a series of reading time studies reported by Delogu et al. (2010). They examined eye-movements in sentences like “*The beer was warm*” following non-actual contexts introduced by intensional verbs (“*John wanted a beer*”) vs. actual contexts (“*John drank a beer*”), but failed to observe any cost at the definite NP or the following verb. This null effect might actually be the result of two competing effects, both originating from the non-actual context: a cost

² Dwivedi et al. (2006) also report a significant effect (a late negativity starting at about 500 ms) on the pre-target region (the pronoun). As previously noted (Delogu, Vespignani, & Sanford, 2010), the pre-stimulus baseline for the target region might have picked up this effect and created an ERP artifact on the target (see Steinhauer & Drury, 2012).

due to the difficulty of establishing an antecedent for the definite NP, combined with a facilitation resulting from the NAI triggered by the intensional verb, which sets up a QUD about whether or not John got a beer. Any discourse continuation that is interpreted as initiating an answer to - or a comment on - this QUD (e.g., *The beer that he drank*) is predicted to be easier to process.

The present experiment used ERPs to rule out the possibility that the failure to observe the processing cost expected in Delogu et al. was due to the methodology used in their studies. While with eye-movements participants are free to move their eyes throughout the text, so that words can be skipped or benefit from parafoveal preview, the rapid serial visual presentation (RSVP) technique (Foster, 1970) used in ERP studies forces each word to be read for a fixed amount of time. ERPs, thus, reflect the incremental processing of each word in a sentence, with no influence from the right of the fixated word.

We therefore tested sentence pairs similar to those used in Delogu et al. (2010) – as in (1) – to measure the ERP responses to definite NPs following actual (1a) vs. non-actual (1b) contexts.

- (1) a. *Giorgio bevve una birra dopo la passeggiata. La birra era calda.*
 (Giorgio drank a beer after the walk. The beer was warm.)
- b. *Giorgio voleva una birra dopo la passeggiata. La birra era calda.*
 (Giorgio wanted a beer after the walk. The beer was warm.)

We hypothesize that the NAI raised in non-actual contexts sets up a QUD denoted by two alternative states of affair (*{John gets a beer; John does not get a beer}*), thereby raising the expectation that the following discourse will select one of these alternatives. No such alternatives are introduced in actual contexts, where several different coherent continuations are possible. Based on previous findings (Dwivedi et al., 2006), definite NPs in non-actual contexts should elicit a cost reflecting more

difficult reference resolution processes, with no influence from the implicit QUD.³

Alternatively, if comprehenders are sensitive to the QUD introduced by the NAI, there should be a facilitation for definite NPs in non-actual contexts, to the extent that the definite is interpreted as being relevant to the QUD (e.g., because it refers to the object of the desired state).

Participants

Twenty-seven right-handed students and staff at the University of Trento (15 females, mean age 26.9) gave written informed consent and took part in the experiment. They were all native Italian speakers and had a normal or corrected-to-normal vision.

Materials

A set of 60 target sentences in Italian was created, each of which was coupled with two context-sentences which differed only for the verb. One version contained control verbs like *acquistò* (bought) and *bevve* (drank), establishing an Actual context. The other half contained intensional verbs like *voleva* (wanted) and *cercava* (sought) considered to raise a non-actuality implicature (Non-Actual context). The verbs were followed by an indefinite NP lexically identical to the definite NP at the beginning of the target sentence. The verb of the target sentence was always in the past tense, indicative mood. The materials were adapted from Delogu et al. (2010), Experiment 1 (see example (1) above).

The 120 sentence pairs were arranged into two lists, according to a Latin square design. Experimental items were combined with 160 filler sentences of various type and length.

³ Based on Dwivedi et al. (2006)'s findings on the pronoun, an increase in cognitive load on the definite should be observed even in the case of anaphoric reference in the context of modal subordination (i.e., if the anaphor is initially interpreted as occurring in a non-factual clause). As shown by Dwivedi et al. (2006), anaphoric processes in modal contexts are more demanding, presumably because modal contexts are semantically and structurally more complex than factual contexts.

Procedure

Each participant was seated in front of a computer screen. Participants silently read the sentence pairs presented word-by-word at the center of the computer screen, while minimizing eye-movements and blinks. After a fixation mark at the beginning of each sentence, words appeared for 300 ms followed by a blank screen interval of 300 ms, except for the last word of the context and the target sentence where the blank screen interval lasted 1000 ms. Comprehension questions requiring a yes/no answer followed 25% of the trials. At the fixation mark, participants could blink and proceed to the next sentence.

Electroencephalographic (EEG) recording

The EEG was recorded from 25 scalp electrodes positioned according to the 10-20 system (Sharbrough et al., 1991). Eye movements and blinks were monitored by means of four additional electrodes placed beneath the left and right eye and at external canthi. All sites were referenced to the left mastoid and the ground was in the Afz site. EEG was amplified in DC with a low-pass filter at 250 Hz and digitized at 250 Hz. Impedance was kept below 5 kOhm for all EEG sites.

Data analysis

Continuous EEG was re-referenced to the average mastoids activity. Epochs around the target word (the noun at the beginning of the second sentence), lasting from 200 ms before onset to 1200 ms, were extracted and visually inspected in order to detect artifacts. Four participants were excluded due to excessive artifacts. For the remaining 23 participants, 5.3% of epochs were rejected; the number of rejections did not differ across conditions ($F < 1$). After a 200 ms pre-stimulus baseline, single subject average ERPs were computed over artifact-free trials for each critical word and condition. We conducted ANOVAs with Actuality (Actual, Non-Actual) as within-subjects factor over midline and lateral sites. The analyses of midline sites (Fz, Cz, Pz) included longitude (anterior, central, posterior) as an additional within-subjects factor. Data from lateral

sites were grouped into four regions of interest (ROIs): Left-Anterior (FC5, F3, F7); Right-Anterior (FC6, F4, F8); Left-Posterior (CP5, P3, P7); Right-Posterior (CP6, P4, P8). In addition to Actuality, ANOVAs over lateral sites included Hemisphere (left, right) and Longitude (anterior, posterior) as within-subjects factors.

Results and Discussion

Grand-average ERPs time-locked to the complement nouns at electrode Cz are shown in Figure 1.

Visual inspection of the waveforms shows only a broadly distributed negativity in the N400 time-window, with nouns in the Non-Actual condition showing a reduced negativity compared to nouns in the Actual condition. The ANOVA between 250-450 ms time-window for the midline electrodes showed a significant effect of Actuality, $F(1, 22) = 5.15$, $p < .05$, and no interaction with Longitude. The analysis of the ROIs showed an effect of Actuality, $F(1, 22) = 4.86$, $p < .05$, and no interaction with Longitude or Hemisphere.

In sum, we observed a reduced N400 on *birra* (beer) following a non actual context (*wanted a beer*) vs. a control context (*drank a beer*), and no P600 effect. This result is consistent with the findings of Delogu et al. (2010) and suggests that processing the definite NP was facilitated when the noun could be interpreted as relevant to addressing the implicit QUD (*Did John get a beer?*). It could be argued, however, that rather than a facilitation for the non-actual context, the N400 effect at the noun reflects a “repeated NP penalty” (e.g., Almor, 1999) for the definite NP in the actual condition, where a pronoun would have been more felicitous (see e.g., Ledoux, Gordon, Camblin, & Swaab, 2007; Swaab, Camblin, & Gordon, 2004, for repeated-*name* penalty N400 effects). According to theories that attempt to explain distributional patterns of referential forms in language use (e.g., Almor, 1999, 2000; Almor & Eimas, 2008; Ariel, 1990; Gordon, Grosz, & Gilliom, 1993; Grosz, Joshi, & Weinstein, 1983), the more accessible a referent is, the less likely it is to be referred to by an explicit referring expression such as a (repeated) definite NP. Consequently, processing a

definite NP referring back to a highly accessible antecedent, as it is the case in actual compared to non-actual contexts, may result in a cost.

To rule out this alternative explanation, Experiment 2 tested sentence pairs in which the definite NP did not involve anaphoric reference to the indefinite in the context sentence, but rather referred to a new discourse entity that, crucially, could or could not be interpreted as initiating a comment on the QUD in a fully acceptable discourse.

Experiment 2

In Experiment 1 we observed a reduced N400 amplitude for definite NPs whose potential antecedents appear in the scope of intensional verbs raising a NAI. While this effect is compatible with the NAI/QUD hypothesis (Clifton & Frazier, 2012), an alternative explanation would attribute the effect to a cost for processing explicit reference to more accessible compared to less accessible antecedents (e.g. Almor, 1999). In Experiment 2, conducted in German, we avoided the repeated noun penalty by testing definite NPs introducing new discourse referents that could or could not be interpreted as addressing the QUD, as established through a norming study. Importantly, to examine whether potential effects of NAI/QUD reflect a general mechanism of discourse processing rather than being beneficial only in cases of imperfect input, we used passages that were judged to be both syntactically and semantically acceptable.

Participants read sentence pairs in four conditions, crossing the factors type of context (Actual, Non-Actual) and type of target (Commenting, Non-commenting). An example of the experimental items is given in (2):

- (2) a. Non-Actual context / commenting target:

Peter hatte einen langen Tag und wollte ein Bier. Die Kneipe war bis Mitternacht geöffnet.

(Peter had a long day and wanted a beer. The bar was open till midnight.)

- b. Non-Actual context / non-commenting target:

Peter hatte einen langen Tag und wollte ein Bier. Das Essen war bereits auf

dem Tisch.

(Peter had a long day and wanted a beer. The meal was already on the table.)

- c. Actual context / commenting target:

Peter hatte einen langen Tag und trank ein Bier. Die Kneipe war bis Mitternacht geöffnet.

(Peter had a long day and drank a beer. The bar was open till midnight.)

- d. Actual context / non-commenting target:

Peter hatte einen langen Tag und trank ein Bier. Das Essen war bereits auf dem Tisch.

(Peter had a long day and drank a beer. The meal was already on the table.)

According to the NAI/QUD hypothesis, the non-actual contexts (2a) and (2b) should raise a specific QUD, namely whether or not Peter got a beer, thereby generating the expectation that the following discourse will address this QUD. No specific expectation should arise with actual contexts (2c) and (2d), as no specific QUD involving alternatives is introduced in such contexts (Clifton & Frazier, 2012, 2017). Following Clifton and Frazier (2017), if comprehenders are sensitive to QUDs in fully acceptable discourses, and interpret each chunk of the upcoming discourse in the context of the accepted QUD (Clifton & Frazier, 2017), we should observe a facilitation for processing definite NPs that are interpreted as initiating a comment on the QUD - like *Die Kneipe* (The bar) in (2a) - and/or a cost for processing definite NPs that are interpreted as failing to address the QUD - like *Das Essen* (The meal) in (2b).

Participants

Thirty right-handed students at Saarland University (23 females, mean age: 23.5) gave written informed consent and took part in the ERP experiment. They were all native speakers of German and had normal or corrected-to-normal vision.

Materials

Similar to Experiment 1, we created 100 sentence pairs in German in which the context-sentences differed only for the verb: one version contained control verbs like *trank* (drank), the other half contained intensional verbs like *wollte* (wanted), which are hypothesized to introduce a QUD (e.g., *Did Peter get a beer?*). The target sentence, however, differed from the previous study so that the NP at the beginning of the target sentence did not involve anaphoric reference to the indefinite NP in the context sentence, but was either commenting on the QUD introduced in the non-actual condition, or not. For example, for the context sentence “*Peter wollte ein Bier*” (Peter wanted a beer), the noun *Kneipe* (bar) was considered initiating a comment on the raised QUD (*Did Peter get a beer?*), whereas the noun *Essen* (meal) was considered to be not commenting on the QUD. The verb of the target sentence was always *war* (was) (see example (2) and the Appendix).

The nouns in the commenting vs. non-commenting conditions did not differ in frequency or length. The average frequency was estimated using the Mannheim German Reference Corpus (DeReKo) corpus (Kupietz, Belica, Keibel, & Witt, 2010). The average frequency class of commenting nouns was 13.23, while that of non-commenting nouns was 12.86. The difference was not significant ($t < 1$). Commenting and non-commenting nouns were on average 7.5 and 7.1 characters long. The difference was not significant ($t < 1.4$).

In order to evaluate whether the chosen NPs were interpreted as initiating vs. not initiating a comment on the QUD as intended, the two types of nouns were pre-tested in combination with the corresponding non-actual context sentence. In a first step, we used a Latin square design to create two lists so that every item appeared in both conditions distributed across the lists. Both lists additionally contained fillers with actual context sentences. Participants were presented with the complete context sentence followed by only the first NP of the second sentence with the task to complete the sentence. Thus, they read, for instance: “Peter had a long day and wanted a beer. The bar...”. Sample completions of what participants actually wrote are listed below:

- “was closed. ”
- “was around the corner.”
- “, however, was already closed.”
- “was closed unfortunately.”

For the version “Peter had a long day and wanted a beer. The meal...” we received completions such as:

- “consisted of fried potatoes.”
- “would be better for him.”
- “was spicy/salty.”
- “did not taste well.”

We collected these completions from 24 participants in total with 12 participants per list. All completions were then handed to two annotators with the task to rate whether the completions did or did not comment on the QUD introduced by the context sentence. Completions were to be considered as commenting when they implied the (failure of the) realization of the desired goal state. For example, completions mentioning the bar opening hours, implying that Peter did or did not get a beer there, were rated as ‘commenting’. In contrast, completions that were rather restating or justifying the desired goal state, (e.g., “Peter was thirsty”) were rated as ‘NOT commenting’.

The annotators showed a moderate inter-annotator agreement (Cohen’s Kappa = 0.46). Based on their annotations, we excluded 32 items for which less than 60% of the completions behaved as predicted. More precisely, only items for which the supposedly commenting vs. non-commenting sentences were indeed considered by the annotators as commenting vs. non-commenting on the implicit QUD were included in the main part of the experiment. Table 1 depicts a bad item with three sample completions per condition. In the non-commenting condition (*Die Ehefrau*), completions were rated as not commenting on the goal state of finding an apartment, as expected. In the commenting condition (*Die Maklerin*), however, completions were

expected to comment on the goal state, but annotators rated them as not commenting (NC in column “Rating”). We therefore discarded this item. Table 2 shows an example of a good item, where the conditions behaved as expected.

To assess whether the remaining 68 items formed fully acceptable discourses, we asked 48 participants that did not take part into the EEG experiment to rate on a 7-point Likert scale how well and plausibly the second sentence continued the first one (1 = entirely incoherent - 7 entirely coherent). We also created a further control condition in which the context sentences in the actual and non-actual conditions were paired with unrelated continuations. As expected, the unrelated pairs were judged as generally incoherent (Actual-Unrelated: 1.55; Non-Actual-Unrelated: 1.53), while the four experimental conditions were judged as broadly coherent (Actual/Commenting: 6.16; Actual/Non-commenting: 5.84; Non-Actual/Commenting: 6.02; Non-Actual/Non-commenting: 5.53). We used cumulative link mixed models (Christensen, 2018) to assess whether Actual and Non-Actual contexts differed within the Commenting and Non-Commenting conditions respectively. Interestingly, while within the Commenting condition, Actual and Non-actual contexts did not differ, $\beta = -0.19$, $SE = 0.13$, $z = -1.51$, $p = .13$, within the Non-Commenting conditions, Non-Actual items were judged as less coherent than Actual items, $\beta = -0.49$, $SE = 0.12$, $z = -4.15$, $p < .001$. The penalty for discourses in which the continuation sentence does not comment on an implicit QUD is consistent with the predictions of the NAI/QUD hypothesis, namely that a cost should be engendered when a specific QUD raised in the context is not addressed. Importantly, the fact that, overall, the items were rated as coherent (the mean ratings of all four experimental conditions were 5.89 points) indicates that the materials used in the current experiment formed fully acceptable discourses.

The 68 items were arranged into four lists, according to a Latin square design. Experimental items were combined with 68 filler sentences of various type and length. Each list was pseudo-randomized. In order to counterbalance for order of presentation, we also prepared mirrored versions of the lists which presented the items in opposite order.

Procedure

The procedure was identical as in Experiment 1, except for the duration of the inter-word blank screen interval, which was 200 ms.

EEG recording

The EEG recording was similar to the one used in Experiment 1. The EEG was digitized at 500 Hz.

Data analysis

Data analysis was similar as in Experiment 1. The EEG signal was band-pass filtered offline at 0.01-40 Hz in order to attenuate skin potentials and other low voltage changes. Five participants were rejected due to excessive artifacts. Critical words were the first noun in the second sentence (the “target” noun) and the discourse-final word (DFW), the last word of the second sentence. This latter region was considered to assess whether potentially commenting vs. non-commenting NPs did indeed lead into a sentence that then contained a comment on the QUD (or not) – at a point in which comprehenders were finally able to verify this. Thus, DFWs served to confirm whether the NP managed to foreshadow what was in fact conveyed by the whole sentence.

We computed mean amplitudes for the target nouns in each condition in the 350-550 ms (N400) time-window⁴ and in the 600-800 ms (P600) time-window. For the DFW target, we computed mean amplitudes on a later time-window, between 800-1000 ms, where sentence final wrap-up effects have been attested (e.g., Osterhout & Holcomb, 1992). Within each time window, ANOVAs were computed with Actuality (Actual, Non-Actual) and Comment (Commenting, Non-Commenting) as within-subjects factors. The topographic distribution of the effects was investigated using the same ROIs as in Experiment 1. Planned simple effects ANOVAs comparing commenting targets in

⁴ This time-window was chosen based on visual inspection of the waveforms (see Fig. 2). In Experiment 1, however, the N400 effect appeared in an earlier time-window, between 250 and 450 ms. We think that this difference in latency may depend in part on the distinct stimulus onset asynchronies (SOAs) employed in the two experiments: in Experiment 1 words appeared with an SOA of 600 ms, while in Experiment 2 they appeared with an SOA of 500 ms. With faster presentation rates, both the onset and peak of the N400 effect are delayed (see, for example, Dambacher et al., 2012)

actual vs. non-actual contexts and non-commenting targets in actual vs. non-actual contexts were computed following significant Actuality x Comment interactions. The Greenhouse-Geisser correction was applied whenever there were more than one degree of freedom in the numerator. In such cases, the corrected p -value is reported.

Results and Discussion

Grand-average ERPs time-locked to the target noun (electrode P3) and to the DFW (electrode Cz) are shown in Figure 2 and Figure 4, respectively.

Visual inspection of the waveforms at the target noun shows a larger negativity in the N400 time-window for non-commenting nouns in non-actual compared to actual contexts. The effect appears to be larger over left electrodes. The ANOVA between 350-550 ms time-window for the midline electrodes did not show any significant effect (all F s < 1.1). The analysis of lateral electrodes, however, revealed an Actuality x Comment x Hemisphere interaction, $F(1, 24) = 9.54, p < .01$.

To investigate this interaction further, we split the data by hemisphere based on the main effect of this factor, $F(1, 24) = 24.78, p < .01$. An ANOVA performed on the two subsets revealed an interaction of Actuality and Comment only on the left hemisphere, $F(1, 24) = 7.05, p < .05$. The comparison between the two non-commenting (*meal*) conditions showed an effect of Actuality, $F(1, 24) = 6.41, p < .05$, with a larger negativity for the non-actual (*wanted*; mean = $-0.21 \mu\text{V}$) condition compared to the actual (*drank*; mean = $0.94 \mu\text{V}$) condition. The comparison between the two commenting conditions (*bar*) showed no significant effects. Additionally, we analyzed the contrast between commenting and non-commenting nouns in the two types of contexts. The comparison between the two nouns in the non-actual condition showed an effect of Comment, $F(1, 24) = 11.69, p < .01$, with a larger negativity for non-commenting (*meal*; mean = $-0.21 \mu\text{V}$) compared to commenting (*meal*; mean = $1.24 \mu\text{V}$) nouns. The ANOVA for the actual condition showed no significant effects ($F < 2$). The interaction between the four conditions in the N400 time-window is shown in Figure 3.

In the time-window between 600 and 800 ms, the ANOVA for midline electrodes did not show any significant effect (all $F_s < 1.1$). The analysis of lateral electrodes showed an Actuality x Comment x Hemisphere x Longitude interaction, $F(1, 24) = 4.71$, $p < .05$. The comparison between non-commenting nouns (*meal*) revealed a marginal Actuality x Hemisphere x Longitude interaction, $F(1, 24) = 3.57$, $p < .07$, with non-commenting nouns being more negative following non-actual compared to actual contexts. This marginal effect suggests that the negativity elicited in non-actual contexts in the N400 time-window was long-lasting (see Fig. 2). The comparison between actual and non-actual contexts for commenting nouns (*bar*) also showed a marginal Actuality x Hemisphere x Longitude interaction, $F(1, 24) = 3.68$, $p < .07$, suggesting that commenting nouns were more positive following non-actual contexts compared to actual contexts over left-posterior electrodes. This marginal positivity might reflect more demanding discourse updating processes (Brouwer et al., 2012). More specifically, previous work has shown that partial answers to constraining questions elicit a P600 effect relative to a control condition (Hoeks, Stowe, Hendriks, & Brouwer, 2013). The positivity was interpreted as indexing the effort involved in integrating pragmatic inferences into the mental representation of the unfolding dialogue. In a similar vein, the nouns addressing the QUD may be perceived as partial answers to the current QUD, with the positivity indexing the effort involved in computing and integrating the two alternatives that potentially answer the QUD (e.g., *John did* vs. *did not drink a beer in the bar*). However, since the effect was only marginally significant, we will not further discuss it.

At the DFW, the ANOVA over midline electrodes in the 800-1000 ms time-window showed a marginal larger negativity for the non-actual non-commenting condition compared to the actual non-commenting condition, $F(1, 24) = 3.79$, $p < .07$, and no interaction with Longitude. No differences were found between the non-actual and actual commenting conditions ($F < 1$). Over lateral electrodes, the non-actual non-commenting condition showed a larger negativity than the actual non-commenting condition, $F(1, 24) = 6.62$, $p < .01$. The comparison between the actual and non-actual

commenting condition did not show significant effects ($F_s < 2.1$). No interactions with Longitude or Hemisphere were observed.

To summarize, while in Experiment 1 we found a facilitation for (repeated) definite NPs that can be understood as initiating a comment on an implicit QUD, in Experiment 2 we observed a clear cost (a N400-like effect) for discourse continuations that do *not* address the QUD, a result that is consistent with the NAI/QUD hypothesis (Clifton & Frazier, 2012, 2017; Grant et al., 2012). This effect resembled the standard N400 effect in terms of its morphology and time-course, but not entirely in terms of its distribution. The standard N400 effect has a centro-parietal distribution, sometimes slightly larger over the right sites for written words (Kutas & Federmeier, 2011), whereas the effect observed in this experiment was significant over the left hemisphere. While the reason for this difference is not clear, we can at least infer that the observed negativity recruits some (even though not all) of the neural generators that underlie the standard N400 effect. It should be noticed that, although statistically significant over the whole scalp, also the N400 effect observed in Experiment 1 appears slightly larger over the left sites (see the topographic map in Fig. 1), suggesting that this scalp distribution might be peculiar to the processing of definite NPs in this type of contexts. In any case, at the end of the discourse, where comprehenders could assess whether or not the second sentence was indeed answering the QUD, we observed a larger negativity for sentences that failed to address the QUD introduced in the non actual context, further supporting that non-commenting sentences were indeed more difficult to process compared to their controls.

The comparison between definite NPs that addressed the implicit QUD did not replicate the facilitation effect observed in Experiment 1, suggesting that addressing a QUD is more beneficial when the input is problematic (e.g., when the potential antecedent for the definite NP is inaccessible for definite anaphora, or in the case of mismatch ellipses).

General Discussion

According to the non-actuality implicature hypothesis (Grant et al., 2012), sentences like *John wanted a beer* pragmatically imply the non-actuality of a goal state (*John drinking/having a beer*) and set up a QUD of whether or not the goal state was eventually reached. Previous work (Clifton & Frazier, 2012; Grant et al., 2012) has shown that processing elliptic material that comments on a QUD introduced in a non-actual context is facilitated, especially when the input is ungrammatical (as it is arguably the case for mismatch ellipses). In the present work, we provide electrophysiological evidence that the influence of implicit QUDs extends to fully acceptable discourses.

Experiment 1 showed that a definite NP (e.g., *the beer*) elicits a lower N400 amplitude when its potential antecedent (*a beer*) appears in a non-actual compared to an actual context. Unlike referents appearing in actual contexts, referents introduced in non-actual contexts should be unavailable for anaphoric reference from a factual clause, because the existence of the entity denoted by the expression is not presupposed (e.g., Moltmann, 1997). Processing a definite NP that cannot be linked to its potential antecedent should elicit a cost. This cost, however, is predicted to be attenuated if the definite is interpreted as addressing the implicit QUD introduced in the non-actual condition (e.g., *The beer he eventually drank was ...*). The attenuation of the N400 for definites following a non-actual as opposed to an actual context is compatible with the NAI implicature hypothesis and is consistent with previous findings using the same type of constructions (Delogu et al., 2010) as well as mismatch ellipses (Grant et al., 2012).

It should be noted, however, that in this experiment the target nouns were lexical repetitions of their potential antecedents. Using repeated NP anaphors (rather than pronouns) to refer to more accessible antecedents – as are referents introduced in actual compared to non-actual contexts – may result in a repeated NP penalty (e.g., Gordon et al., 1993; see, however, Almor, 1999; Brennan, 1995; Burkhardt & Roehm, 2007; Gordon et al., 1993, who show that the repeated NP penalty is not observed when the antecedent is in object position). As a matter of fact, using pronouns, Dwivedi et al.

(2006) did not find a facilitation effect, but rather a late negativity at the pronoun in the non-actual condition, interpreted as reflecting an increase in working memory load during an early matching process between the pronoun and its potential antecedent. It is therefore an open question whether the N400 effect observed in Experiment 1 is due to (i) a facilitation for the presence of a NAI introducing a QUD, (ii) a cost for a repeated NP penalty, or (iii) both.

In Experiment 2 we avoided repeated NP anaphors and used instead non-repeated definite NPs that could (e.g., *the bar*) or could not (e.g., *the meal*) be interpreted as initiating a comment on the QUD introduced in the non-actual context, as established through a norming study. Importantly, the two NPs were judged to continue the current discourse in a fully acceptable way. This design allowed us to assess whether 1) definite NPs that fail to address the QUD are more difficult to process compared to a control condition in which no particular QUD is introduced, and therefore no particular expectation about the content of upcoming material is generated (i.e., the actual condition); 2) definite NPs that do address the QUD are less difficult to process compared to a corresponding control condition in which no particular QUD is introduced. The results showed an N400-like effect for definite NPs not addressing the QUD, suggesting that these were less expected and therefore more difficult to process compared to the control condition. This effect was consistent with the ERP responses at the discourse final word, where comprehenders could assess whether the final sentence did or did not provide an answer to the QUD. Here we found a larger sustained negativity for sentences not addressing the QUD compared to the same sentence in the control condition. Enhanced and prolonged negativities at sentence final words have been observed in the presence of semantic or syntactic anomalies (e.g., Ditman, Holcomb, & Kuperberg, 2007; Osterhout & Holcomb, 1992) and interpreted as reflecting a wrap-up process whereby the degree of coherence of the entire sentence or discourse is evaluated (see also Osterhout & Holcomb, 1995; Xiang & Kuperberg, 2015). Similarly, both the wrap-up effect and the offline coherence ratings collected for Experiment 2 jointly indicate that discourses in which a QUD is set up but not

subsequently addressed were perceived as less coherent than discourses in which no specific QUD is introduced, further supporting the NAI/QUD hypothesis. Crucially, the ERPs on the target nouns reveal that QUDs influence processing incrementally at each chunk of the sentence and contribute to the overall coherence of the discourse.

In contrast to Experiment 1, as well as to previous eye-tracking findings (Grant et al., 2012), in Experiment 2 there was no facilitation for definite NPs that do address the QUD. No significant differences were observed in the N400 time-window between commenting nouns in actual vs. non actual contexts. Rather, in this time-window we found that commenting and non-commenting nouns were equally easy to process when no particular QUD was at stake, but they elicited effects when the context introduced a QUD, with nouns failing to address it being more negative than nouns addressing it. This clearly indicates that, while the two NPs were equally expected in the absence of a specific QUD, the presence of a QUD raised the expectation that upcoming material would address it.

To summarize, our results suggest that the presence of a QUD helps the processing of materials that appear in the context of problematic input, as it is the case for ellipsis in which the elided phrase mismatches syntactically with its antecedent (Grant et al., 2012), or for anaphoric definite NPs whose potential antecedents are inaccessible for definite anaphora (when compared to definite NPs possibly giving rise to a repeated NP penalty, as in Experiment 1). In the context of fully acceptable discourse, the presence of a QUD disrupts processing of discourse continuations that *fail* to address it, a result that is entirely consistent with previous eye-tracking findings using non-mismatching ellipsis (Clifton & Frazier, 2012).

Our results are consistent with other approaches to narrative comprehension, such as the event-indexing model (e.g., Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995) and other situation model accounts (e.g., Bower & Morrow, 1990; Johnson-Laird, 1983; van Dijk & Kintsch, 1983), in which characters' goals play an important role in constructing a coherent mental representation of the situation described. According to these approaches, readers tend to interpret incoming

information as relevant to the protagonist's goal by inferring the connections between these goals and other elements of the story (e.g., Bower, 1982; Bower & Rinck, 1999; Egidi & Gerrig, 2006; Huitema, Dopkins, Klin, & Myers, 1993). Our results suggest that, at least when a non-actuality implicature is triggered, readers expect incoming information to be specifically related to the *achievement* of the goal (i.e., to the QUD) rather than more generically to the goal itself. If the latter was the case, no processing cost should be observed on "meal" vs. "bar" in our example, since some property of the meal (e.g., being salty) could be mentioned to provide a reason for the character's goal (i.e., getting a beer). Thus, while our findings are consistent with previous literature on discourse processing in showing that readers are sensitive to characters' goals, they further indicate that, under certain circumstances, readers expect more specific information about that goal, namely an answer to the QUD.

Overall, our results provide further evidence that NAIs are computed online not only from modal verbs, as shown in previous studies (Clifton & Frazier, 2012; Grant et al., 2012), but also from so called *intensional* verbs like *want*, *need*, or *hope for* (as hypothesized by Delogu et al., 2010 and Grant et al., 2012). Our findings suggest that these verbs set up a QUD of whether or not the goal-state that is wanted, needed, or hoped for will eventually be realized, thereby playing an important role in generating expectations about upcoming material. When discourse continuations address the implicit QUD, comprehension proceeds with no particular effort. By contrast, when the upcoming material fails to comment on the QUD, the discourse is perceived as less coherent and processing is more demanding. Importantly, our results indicate that comprehenders do not wait until the end of a sentence to relate it to an already accepted QUD but, as previously argued (Clifton & Frazier, 2012, 2017), the effects of a QUD can be seen during the ongoing processing of the sentential material. Thus, the present study provides evidence that QUDs are an important factor in guiding discourse comprehension and should therefore be more closely addressed in future research.

Acknowledgments

This work was supported by the Cluster of Excellence “Multimodal Computing and Interaction” of the German Excellence Initiative, and the SFB/CRC 1102 "Information density and linguistic encoding" awarded by the Deutsche Forschungsgemeinschaft (DFG). Further support was provided by the Spanish Ministry of Economy and Competitiveness (MINECO), the Agencia Estatal de Investigacion (AEI), the Fondo Europeo de Desarrollo Regional (FEDER) (grant PSI2015-65694- P, ‘Severo Ochoa’ programme SEV-2015-490 for Centres of Excellence in R&D), and the Basque Government (grant PI_2016_1_0014). We thank Matthew Crocker, Les Sikos, and Lyn Frazier for comments on previous versions of this manuscript.

Disclosure statement

The authors report no conflict of interest.

References

- Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, *106*, 748-765.
- Almor, A. (2000). Constraints and mechanisms in theories of anaphor processing. In M. Pickering, C. Clifton, & M. Crocker (Eds.), *Architectures and mechanisms for language processing*. England: Cambridge University Press.
- Almor, A., & Eimas, P. D. (2008). Focus and noun phrase anaphors in spoken language comprehension. *Language and Cognitive Processes*, *23*(2), 201-225.
- Ariel, M. (1990). *Accessing noun-phrase antecedents*. London and New York: Routledge.
- Benz, A., & Jasinskaja, K. (2017). Questions Under Discussion: From sentence to discourse. *Discourse Processes*, *54*(3), 177-186.
- Bower, G. H. (1982). Plans and goals in understanding episodes. *Advances in Psychology*, *8*, 2-15.
- Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science*, *247*, 44-48.
- Bower, G. H., & Rinck, M. (1999). Goals as generators of activation in narrative understanding. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension causality and coherence: Essays in honor of Tom Trabasso* (p. 111-134). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, *10*(2), 137-167.
- Brouwer, H., Crocker, M. W., Venhizen, N. J., & Hoeks, C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, *41*(Suppl Suppl 6), 1318-1352.
- Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127-143.
- Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms:

- Evidence from event-related brain potentials. *Brain and Language*, *98*, 159-168.
- Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *NeuroReport*, *18*(17), 1851-1854.
- Burkhardt, P., & Roehm, D. (2007). Differential effects of saliency: An event-related brain potential study. *Neuroscience Letters*, *413*, 115-120.
- Carlson, L. (1983). *Dialogue games: An approach to discourse analysis*. Dordrecht: Reidel.
- Christensen, R. H. (2018). *ordinal – regression models for ordinal data*. (R package version 2018.8-25. <http://www.cran.r-project.org/package=ordinal>)
- Clifton, C. J., & Frazier, L. (2012). Discourse integration guided by the ‘question under discussion’. *Cognitive Psychology*, *65*(2), 352–379.
- Clifton, C. J., & Frazier, L. (2017). Context effects in discourse: The question under discussion. *Discourse Processes*, *55*(2), 105-112.
- Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012). Stimulus onset asynchrony and the timeline of word recognition: event-related potentials during sentence reading. *Neuropsychologia*, *50*(8), 1852-70.
- Delogu, F., Vespignani, F., & Sanford, A. J. (2010). Effects of intensionality on sentence and discourse processing: Evidence from eye-movements. *Journal of Memory and Language*, *62*(4), 352–379.
- Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology*, *44*(6), 927-935.
- Dwivedi, V. D., Phillips, N. A., Lague-Beauvais, M., & Baum, S. R. (2006). An electrophysiological study of mood, modal context, and anaphora. *Brain Research*, *1117*(1), 135–153.
- Egidi, G., & Gerrig, R. J. (2006). Readers’ experiences of characters’ goals and actions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(6), 1322-1329.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *Journal of Memory and Language*, *8*(3),

373-392.

- Foster, K. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Percept*, 8, 215-221.
- Frazier, L. (2008). Processing ellipsis: A processing solution to the undergeneration problem. In *Proceedings of the 26th west coast conference on formal linguistics* (pp. 21-32).
- Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17, 311-347.
- Grant, M., Clifton, C., & Frazier, L. (2012). The role of non-actuality implicatures in processing elided constituents. *Journal of memory and language*, 66(1), 326-343.
- Grosz, B. J., Joshi, A. K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. In *Proceedings of 21st annual meeting of the association of computational linguistics* (p. 44-50).
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439-483.
- Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, 10, 41-53.
- Hobbs, J. R. (1985). On the coherence and structure of discourse. In L. Polanyi (Ed.), *The structure of discourse*. Norwood, N. J.: Ablex.
- Hoeks, J. C., Stowe, L. A., Hendriks, P., & Brouwer, H. (2013). Questions left unanswered: How the brain responds to missing information. *PLoS ONE*, 8(10), e73594.
- Huitema, J. S., Dopkins, S., Klin, C., & Myers, J. (1993). Connecting goals and actions during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1053-1060.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.
- Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, 1,

3-44.

- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publication.
- Kehler, A., & Rohde, H. (2017). Evaluating an expectation-driven QUD model of discourse interpretation. *Discourse Processes*, *54*(3), 219-238.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363-394.
- Kupietz, M., Belica, C., Keibel, H., & Witt, A. (2010). The German Reference Corpus DeReKo: A primordial sample for linguistic research. In N. Calzolari et al. (Eds.), *Proceedings of the 7th conference on international language resources and evaluation (lrec)*. Valletta, Malta: European Language Resources Association (ELRA).
- Kuppevelt, J. V. (1995). Discourse structure, topicality and questioning. *Journal of Linguistics*, *31*(§), 109-147.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161-163.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920-933.
- Ledoux, K., Gordon, P. C., Camblin, C. C., & Swaab, T. Y. (2007). Coreference and lexical repetition: Mechanisms of discourse integration. *Memory and Cognition*, *35*(4), 801-915.
- Moltmann, F. (1997). Intensional verbs and quantifiers. *Natural Language Semantics*, *5*, 1-52.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785-806.
- Osterhout, L., & Holcomb, P. J. (1995). Event-related potentials and language

- comprehension. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiological studies of human cognition*. Oxford: Oxford University Press.
- Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics*, 49, 91-136.
- Roberts, C. (2004). Discourse context in dynamic interpretation. In L. Horn & G. Ward (Eds.), *Handbook of contemporary pragmatic theory* (p. 197-220). Oxford: Blackwell.
- Sanders, T. J. M., & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29(1), 37-60.
- Sharbrough, F., Chatrian, G.-E., Lesser, R. P., Lüders, H., Nuwer, M., & Picton, T. W. (1991). American electroencephalographic society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 8, 200-202.
- Spotorno, N., Cheylus, A., Henst, J.-B. V. D., & Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PLoS ONE*, 8(6), e66839.
- Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, 120, 135-162.
- 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(5), 715-726.
- van den Broek, P. (1990). Causal inferences and the comprehension of narrative texts. In A. C. Graesser & G. H. Bower (Eds.), *Psychology of learning and motivation* (Vol. 25, p. 175-196). New York: Academic Press.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Van Geenhoven, V., & McNally, L. (2005). On the property analysis of opaque complements. *Lingua*, 115, 885-914.
- Xiang, M., & Kuperberg, G. R. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 30(6), 648-672.
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation

models in narrative comprehension: An Event-Indexing Model. *Psychological Science*, 6(5), 292-297.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 386-397.

Appendix

The following table lists a sample of the experimental items used in Experiment 2. The first line provides the context sentence with both the intensional and control verbs. The next two lines show the two target sentence types, i.e., the commenting sentence (CO) and the non-commenting sentence (NC). A translation of all sentences is provided in brackets. The percentages following the target sentences display the annotators' ratings associated with the respective sentences, i.e., how often the completions were rated as commenting (in case of a CO sentence) or non-commenting (in case of a NC sentence) respectively.

Item	Context sentence	Ratings
1	Karl machte Lasagne und kochte/brauchte eine Bechamelsauce. (Karl prepared a lasagna and cooked/needed a bechamel sauce.) CO: Das Rezept war dabei sehr hilfreich. (The recipe was very helpful.) NC: Das Hackfleisch war bereits angebraten. (The minced meat was already seared.)	64% 93%
2	Klara hatte einen anstrengenden Vormittag und machte/brauchte ein Nickerchen. (Klara had a long exhausting morning and took/needed a nap.) CO: Das Sofa war sehr bequem. (The sofa was very comfortable.) NC: Die Arbeit war sehr ermattend. (The work was very tiring.)	64% 86%
3	Sarah hatte einen Termin und nahm/benötigte ein Taxi. (Sarah had an appointment and took/needed a taxi.) CO: Der Fahrer war zum Glück sehr pünktlich. (The driver, luckily, was on time.) NC: Der Doktor war schon ungeduldig. (The doctor was already waiting impatiently.)	93% 86%
4	Max liebte Tiere und streichelte/wollte eine Katze. (Max loved animals and stroked/wanted a cat.) CO: Die Allergie machte ihm allerdings schwer zu schaffen. (The allergy, however, was giving him a hard time.) NC: Die Gesellschaft machte ihm große Freude. (The company made him very happy.)	93% 71%
5	Felix trainierte für einen Triathlon und fuhr/suchte ein Rennrad. (Felix trained for a triathlon and rode/looked for a racing bicycle.) CO: Der Trainer half ihm bei der Auswahl des passenden Rads. (The trainer helped him to find the right bike.) NC: Das Training half ihm beim Erreichen seiner Ziele. (The training helped him to achieve his goals.)	64% 100%
6	Anke fand ein Haar in ihrer Suppe und tadelte/verlangte einen Kellner. (Anke found a hair in her soup and reprimanded/asked for a waiter.) CO: Die Bedienung entschuldigte sich für die Unannehmlichkeiten. (The server apologised for the inconvenience.) NC: Die Begleitung entschuldigte sich später für den Ton. (The company later apologised for the harshness.)	100% 100%
7	Simone fuhr oft zu Außenterminen und borgte/forderte einen Dienstwagen. (Simone often drove to outside appointment and borrowed/demanded a company car.) CO: Der Schlüssel wurde ihr von ihrem Chef ausgehändigt. (The key was handed to her by her boss.) NC: Die Anfahrt wurde damit viel einfacher. (The drive with it became much easier.)	86% 71%
8	Fabian hatte bald ein Vorstellungsgespräch und probierte/suchte eine Hose. (Fabian had a job interview and tried/looked for a pair of trousers.) CO: Die Schneiderei machte ihm einen guten Preis. (The tailor made him a good price.) NC: Die Aufregung machte ihm weiche Knie. (The agitation made him go weak at the knees.)	79% 86%
9	Florian war erkältet und benutzte/brauchte ein Taschentuch. (Florian had a cold and used/needed a handkerchief.) CO: Die Packung war schon fast leer. (The box was already almost empty.) NC: Die Nase war vom vielen Schnäutzen schon ganz entzündet. (The nose was already very red from all the nose blowing.)	64% 83%
10	Alexandra war sehr müde und trank/forderte einen Kaffee. (Alexandra was very tired and drank/demanded a coffee.) CO: Die Tasse war sehr heiß. (The mug was very hot.) NC: Das Sofa war sehr bequem. (The sofa was very comfortable.)	100% 71%
11	Luisa saßen ganzen Tag in ihrer Wohnung und machte/brauchte einen Spaziergang. (Luisa was sitting in her apartment all day and went for/needed a walk.) CO: Der Rundgang war eine willkommene Abwechslung. (The walkabout was a nice change.)	79%

	NC: <i>Das Wetter war allerdings nicht besonders schön.</i> (The weather, however, wasn't very nice.)	79%
12	<i>Svenja sah nicht gut und probierte/brauchte eine Brille.</i> (Svenja didn't see very well and tried/needed glasses.)	
	CO: <i>Die Stärke hatte der Optiker festgestellt.</i> (The strength was asserted by the optician.)	79%
	NC: <i>Das Gestell hatte ihr gut gefallen.</i> (The frame was to her liking.)	79%
13	<i>Christoph hatte ein üppiges Essen und trank/benötigte einen Schnaps.</i> (Christoph had an abundant meal and drank/needed a schnapps.)	
	CO: <i>Der Obstbrand schmeckte sehr gut.</i> (The fruit brandy was very delicious.)	93%
	NC: <i>Der Braten schmeckte sehr gut.</i> (The roast was very delicious.)	100%
14	<i>Bastian war sehr bleich geworden und nahm/brauchte ein Sonnenbad.</i> (Bastian had become very pale and took/needed a sunbath.)	
	CO: <i>Die Liege war schon aufgestellt.</i> (The sun lounger was already set up.)	93%
	NC: <i>Der Winter war dieses Jahr sehr düster gewesen.</i> (The winter this year had been very grey.)	100%
15	<i>Kerstin hatte geheiratet und bekam/wollte ein Baby.</i> (Kerstin married and had/wanted a baby.)	
	CO: <i>Die Schwangerschaft war ohne Komplikationen verlaufen.</i> (The pregnancy went through without any troubles.)	64%
	NC: <i>Die Spielsachen waren schon alle gekauft.</i> (The toys were already bought.)	83%

Table 1
Completions and ratings for an item excluded from the experiment (with English translation)

Context sentence

Michael gab das Haus auf und suchte eine Wohnung. (*Michael left the house and was looking for an apartment.*)

Target sentence

Provided NP	Completions	Expected	Rating
Die Maklerin (<i>The agent</i>)	1) verlangte viel Geld. (<i>asked for a lot of money.</i>)	CO	NC
	2) war nervig. (<i>was annoying.</i>)	CO	NC
	3) zeigt ihm eine zwei Zimmerwohnung im Dachgeschoss. (<i>showed him a two-room apartment under the roof.</i>)	CO	NC
Die Ehefrau (<i>The wife</i>)	1) hatte ihn nun endgültig verlassen. (<i>left him for good now.</i>)	NC	NC
	2) half ihm dabei. (<i>was helping him.</i>)	NC	NC
	3) hatte ihn vor die Tuer gesetzt. (<i>kicked him out.</i>)	NC	NC

CO = Commenting; NC = Non-commenting.

Table 2
Completions and ratings for an item included in the experiment (with English translation)

Context sentence				
Peter hatte einen langen Tag und wollte ein Bier. (<i>Peter had a long day and wanted a beer.</i>)				
Target sentence				
Provided NP	Completions	Expected	Rating	
Die Kneipe (The bar)	1) hatte geschlossen. (<i>was closed.</i>)	CO	CO	
	2) hatte aber leider geschlossen. (<i>however, sadly was closed.</i>)	CO	CO	
	3) hatte aber bereits geschlossen. (<i>however, was already closed.</i>)	CO	CO	
Das Essen (The meal)	1) war scharf. (<i>was spicy.</i>)	NC	NC	
	2) war sehr salzig. (<i>was very salty.</i>)	NC	NC	
	3) war kalt. (<i>was cold.</i>)	NC	NC	

CO = Commenting; NC = Non-commenting.

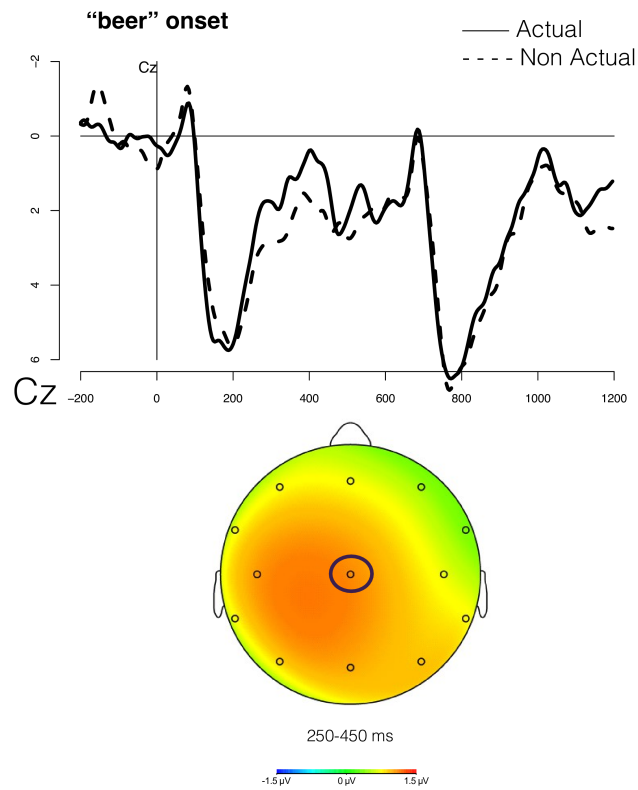


Figure 1. Grand average ERPs time-locked to the onset of the critical noun ("beer") at electrode Cz. Negativity is plotted upwards. The topographic map shows the difference between the Non-Actual and the Actual condition in the N400 time-window.

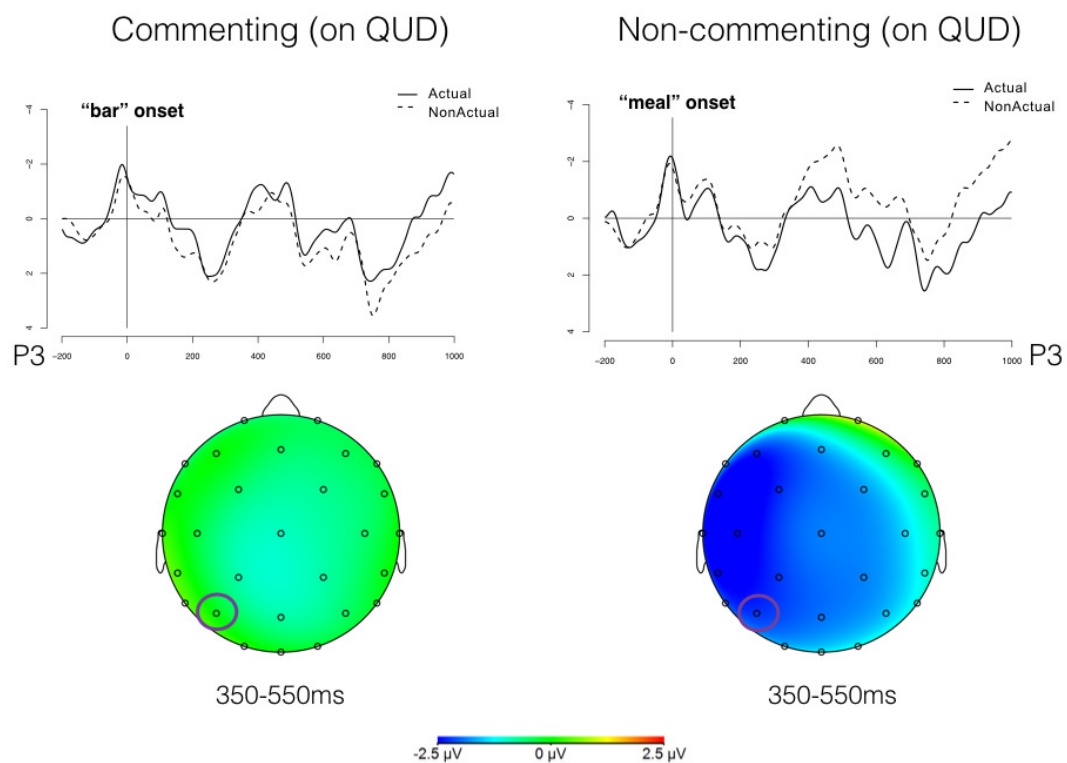


Figure 2. Grand-average ERPs for the target nouns at electrode P3. Negativity is plotted upwards. The left panel shows the ERP waveforms time-locked to commenting nouns ("bar") and the topographic map of the difference between the Non-Actual and the Actual conditions in the N400 time-window. The panel on the right shows the ERP waveforms time-locked to non-commenting nouns ("meal") and the corresponding difference map in the N400 time-window.

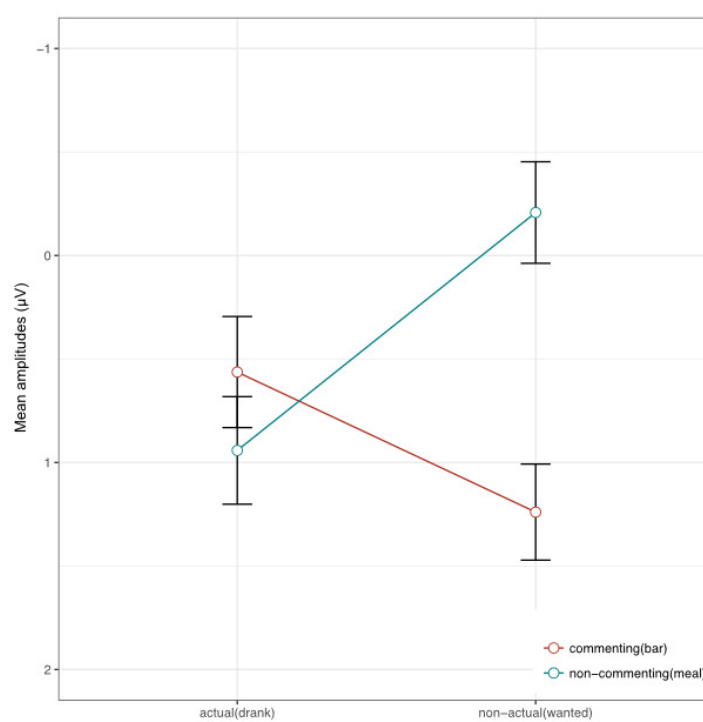


Figure 3. Mean amplitudes in the N400 time-window for the nouns in the four conditions, averaged across electrode sites on the left hemisphere. Negativity is plotted upwards. Error bars indicate SEMs.

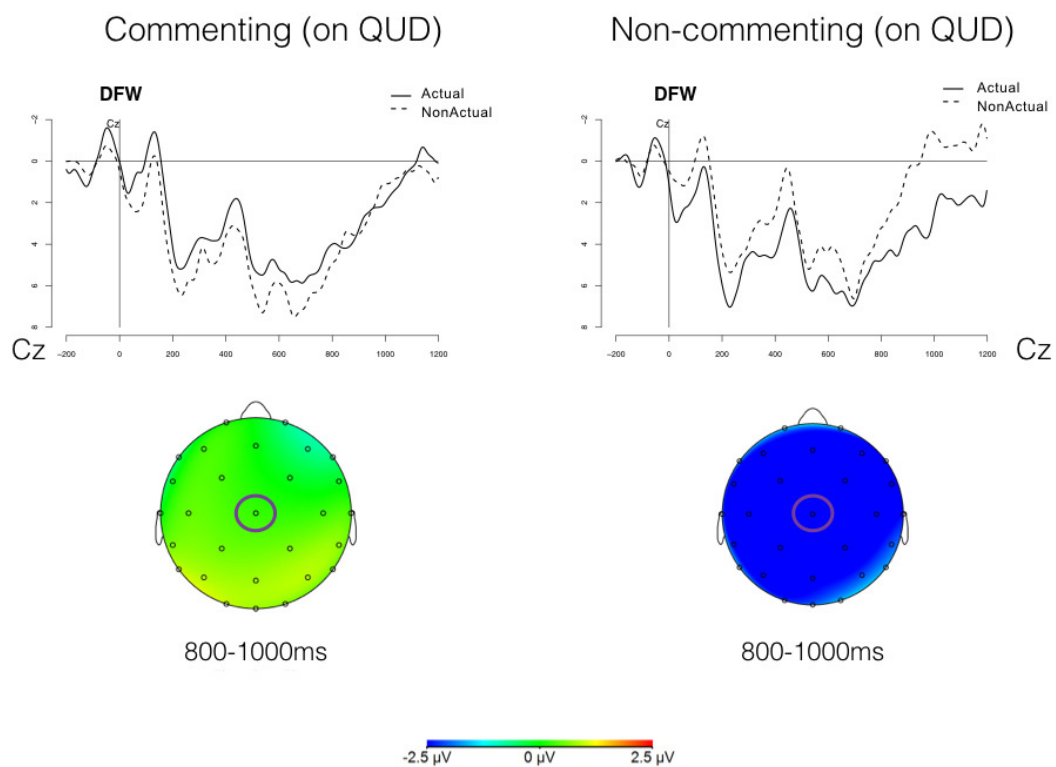


Figure 4. Grand-average ERPs for the discourse final word (DFW) at electrode Cz. Negativity is plotted upwards. The left panel shows the ERP waveforms time-locked to the DFW in the commenting condition and the topographic map of the difference between the Non-Actual and the Actual conditions between 800 and 1000 ms. The panel on the right shows the ERP waveforms time-locked to the DFW in the non-commenting condition and the corresponding difference map between 800 and 1000 ms.