When adding a little is adding too much: how discourse particles force immediate reanalysis and increase processing costs in under-specific contexts

Pia Järnefelt[1] [iD], Gustaf Gredebäck[2] [iD] and Gunnar Norrman[3] [iD]

[1]Department of Scandinavian Languages, Uppsala University, Uppsala, Sweden; [2]Department of Psychology, Uppsala University, Uppsala, Sweden and [3]Centre for Research on Bilingualism, Stockholm University, Stockholm, Sweden
Corresponding author: Pia Järnefelt; Email: pia.jarnefelt@nordiska.uu.se

Abstract
Highly frequent discourse particles (DPs) express speaker attitudes and guide utterance interpretation, but we still lack a satisfactory explanation of how DPs are actually processed. Some results show facilitation, while others show processing costs. Previous studies have aimed to elicit core meanings of DPs embedded in highly plausible contexts, in contrast to more unlikely contexts that force two quite different interpretations. The present study uses a novel eye-tracking experiment where DPs instead are presented in low-constraint contexts. The plausible interpretations consist of two ends of a natural scale: the state change of color that fades or becomes dirty (black to gray or white to gray). This design renders a more direct reflection of how DPs alter context interpretation. Results show that DPs induce immediate reanalysis, and this reanalysis differs in magnitude depending on the kind of DP used. We suggest that the processing of DPs involve three dimensions: i) linguistic intuition about the DP, ii) assumptions about speaker meaning and iii) contextual considerations. The results are interpreted through the communicative principle of language, under-specificity and the maxim of quantity. We also suggest that diverging results from previous studies in the field can be explained using the same analytical lens.

1. Introduction

Successful communication involves a large portion of mind reading. *Can you shut the door?* often means *I need some privacy*, and you might say *It's really late* to indicate to your company that it is time to leave the party or to indicate that your house guests

should leave. When people communicate with each other, they never explicitly spell out all information that an utterance contains. Intonation, prosody, looks and body language, deictic references, visual context and contextual information, as well as world knowledge, all play crucial parts in the interpretation and meaning-making of a conversation as it unfolds. To be able to 'read between the lines' and understand something in context is the ability to pragmatically infer meaning.

In this study, we investigate how pragmatic information is processed and interpreted through an in-depth study of discourse particles (henceforth DPs) that express speaker attitudes. When these particles are introduced into a sentence, the listener is forced to also entertain a pragmatic interpretation of the utterance, in addition to a literal semantic interpretation. DPs can thus facilitate speech processing and comprehension by guiding how the listener should understand what the speaker means.

Language is governed by a communicative principle that leads to the under-specificity of sentences: a speaker will only convey the sufficient amount of information needed in order to be properly understood (Sperber & Wilson, 1986). This is developed from Grice's first maxim of quantity (1989), which states that a speaker should say as much as possible and imply no more. Natural language does not provide full representations of the state of affairs but rather provides information on how to infer meaning (Carston, 1999). The gap between linguistically encoded meaning and what is actually being said needs to be filled by inference and pragmatic principles. This inferential comprehension is ultimately a process where the speaker makes and evaluates hypotheses about meaning and speaker intention based on all contextual information available (visual context, world knowledge, common ground, etc.). This process plays a central role in all human communication (Sperber & Wilson, 2002).

Pragmatics are concerned with that which lies beyond the *propositional content* (i.e., the 'truth value') of sentences. A core tenet of pragmatics is that indirect meaning must be pragmatically inferred and processed in order to be decoded and explicitly understood. A central debate concerns whether speakers process semantic or pragmatic elements first (Parola & Bosco, 2022). Pragmatic inference is claimed to be crucial for successful and efficient use of all resources available in conversations, but common semantic inferences might be automated so that they can override the slow and arduous pragmatic inference processes (Levinson, 2000). An issue with this is that many linguistic elements carry both a semantic and a pragmatic sentiment, as the lines between the two are not clear-cut. The semantic sentiment of an element is the literal one, whereas the pragmatic sentiment is the derived pragmatic interpretation, e.g., *it's fine* could mean 'it is good/excellent' (semantic sentiment), or it could mean 'I am alright with this' (pragmatic sentiment). Semantic/literal meaning has been shown to be computed before the pragmatic meaning in some studies (Bott & Noveck, 2004; Huang & Snedecker, 2009). But other studies have shown that listeners rapidly integrate pragmatic information if it is well-supported by context and the most plausible alternative for decoding the message (Degen & Tanenhaus, 2015). Furthermore, it has been suggested that delays related to the processing of pragmatic inference can be linked to contextual complexity, rather than being intrinsically slower than semantic processing (Grodner et al., 2010).

In line with the communicative principle of only uttering the optimal amount of information, DPs can be used to enrich the content of a sentence so that it can be interpreted more precisely in the proper context, but in many more ways if the context is not clear. At the same time, the presence of a DP can often shorten the sentence, because additional explanations can be omitted. DPs do not alter the

conceptual meaning (e.g., content words), but rather guide the hearer on how they should interpret the sentence and process the conceptual meaning. These qualities are what make DPs *procedural* (Loureda, et al., 2022). DPs anchor sentences in context through indirectly referring to, e.g., visual context, previously stated facts, or shared knowledge. They are a powerful and prolific tool for communication, where one felicitously used DP can convey several sentences worth of information that would otherwise have to be spelled out. A minimal cognitive effort (adding a DP) therefore elicits a maximal cognitive effect (understanding much more than what is literally being said) (Loureda et al., 2022). The present study investigates DPs that can be categorized as belonging to the semantic field of expectation (Aijmer & Simon-Vandenbergen, 2004; but see also van Bergen & Bosker, 2018) and are, in addition to being procedural, carriers of reflexivity. DPs that are reflexive explicitly highlight that the speaker is aware of the fact that communication always takes place in a context. This indicates to the hearer that special notice of such a DP must be made, as it helps guide the interpretation. These metapragmatic particles do not only anchor the utterance in the context but also alter the context (Silverstein, 1976). Although recent years have seen an upsurge in research on DPs and pragmatic online processing, the results diverge.

Based on the empirical findings of a series of eye-tracking studies that used a reading paradigm to investigate DPs in English, Spanish, and German, Loureda et al. (2022) outlined three cognitive principles of discourse marking that predict the effects that introducing a DP has on utterance processing. Adding a DP to a sentence activates basic cognitive processes, such as i) building new access routes to and constraining complex information, ii) optimizing access to assumptions and constraining reanalysis needs and iii) preparing for the integration of upcoming segments. This cognitive activation is due to the inherent procedural meaning of DPs. For instance, DPs such as *sin embargo*, 'however' and *a pesar de ello*, 'despite' in counter-argumentative utterances are claimed to reduce processing costs more than DPs in additive (e.g., *even*) or causal (e.g., *therefore*) utterances. In fact, *all* DPs investigated by Loureda et al. (2022) reduce processing costs compared to utterances without the DPs present, across languages and functional paradigms. These findings also corroborate that introducing a DP into an utterance leads to an immediate modification of the utterance interpretation.

Other studies, however, instead show increased processing costs following the introduction of DPs. Such increases are assumed to reflect greater complexity, introduction of unexpected information or drastic change of meaning (Canestrelli et al., 2013; Filik et al., 2009; Gerwien & Rudka, 2019; van Bergen & Bosker, 2018). For instance, when comparing different focus particles (a subgroup of DPs, henceforth FPs,) in a reading paradigm, Filik et al., (2009) found that *even* had higher processing costs than *only*, due to its higher semantic complexity. In order to isolate the effects of the FPs the study manipulates contexts, using *unlikely* vs *likely* continuations and shows that *even* requires additional processing due to the fact that it conveys unexpected information. The processing patterns for the two Dutch causal connectives *want* and *omdat* (another subgroup of DPs, both meaning approximately 'because') differ because of the interpretational route they invoke. *Want* triggers a subjective causal chain (someone believes *x* is the result of *y*), while *omdat* gives rise to an objective interpretation (*x* happened because of *y*). The subjective connective increases processing costs because the hearer needs to process that it is someone else's beliefs and possibly change perspective because of this. This processing disadvantage

disappears when epistemic markers such as "*Mark thinks that x…*" is introduced (Canestrelli et al., 2013). Van Bergen and Bosker (2018) studied the Dutch interpersonal DPs *eigenlijk* (meaning approximately 'really' or 'actually') and *inderdaad* (meaning approximately 'indeed' or 'really') in a Visual World paradigm (VWP) (Tanenhaus et al., 1995) in order to see whether they would facilitate processing. *Inderdaad* sped up dialogue completion tasks, and *eigenlijk* slowed down dialogue completion tasks in both conditions, *high*-constraint contexts vs *medium*-constraint contexts, but was integrated immediately. A large variability among the participants was also found, and the processing of the adversative DP *eigenlijk* forces the reader to immediately consider a pragmatically inferred unexpected alternative (van Bergen & Bosker, 2018). The same Dutch particles, *eigenlijk* and *inderdaad*, were investigated in an EEG study by Rasenberg et al., (2020) that also aimed at detecting facilitation of processing. The design employed *predictable* and *unpredictable* dialogue endings to investigate the DPs. No effects that could support facilitative processing were detected, but results suggest that the increased processing costs found for *eigenlijk* reflect the fact that this DP forces a pragmatically driven reanalysis of the dialogue continuation. The German FP *sogar* was investigated using both a reading paradigm and a Visual World paradigm, and the presence/absence of the FP was combined with *high* vs *low* expectation of change in dialogue continuations. When *sogar* is used in sentences where there is a high expectation change, participants' eye movements are delayed. This is taken to directly reflect higher processing costs due to a drastic update of the mental representation of the sentence (Gerwien & Rudka, 2019).

Diverging results stem from item and participant variability, and online measures show results that range from processing disadvantages to processing advantages, and upon closer inspection a pattern emerges that might explain the discrepancy. All aforementioned studies manipulate context in combination with the DPs. On the one hand, contexts are designed to exert the most plausible interpretation that the DPs can induce, and it is only in these cases they have been shown to facilitate processing. When the DP fits well with the most plausible interpretation of the whole sentence, i.e., when it is interpreted as having its 'core meaning', it reduces processing costs (Loureda et al., 2022), and when epistemic markers are added that help the pragmatic interpretation, previous disadvantages disappear (Canestrelli et al., 2013). On the other hand, when DPs appear in the unlikely contexts, they increase complexity and thereby induce processing costs (Canestrelli et al., 2013; Filik et al., 2009). Although results on processing costs diverge, there is a strong consensus that participants immediately detect the DPs, which is reflected in various online measures. This immediate integration indicates that DPs are obligatory to take into account once detected. However, hitherto the notion that DPs themselves can cause processing costs in low-constraint contexts has not been sufficiently addressed.

2. The present study

While previous studies were designed with high-constraint contexts to elicit 'core meaning' interpretations of DPs, the present study uses low-constraint contexts that are not intended to induce facilitation *per se,* but rather investigate to what extent DPs impact processing. It combines traditional VWP online measures with offline measures but also introduces pupillometry, in order to gain a deeper understanding of the effect the DP has on its context in incremental online processing.

We tested two Swedish DPs that can induce diametrically different interpretations: on the one hand, the particle *egentligen* ('really', 'actually'), which can have an adversative meaning, and on the other hand *faktiskt* ('actually', 'in fact', 'really'), which carries a confirmatory sentiment.

Consider the following sentences and how the use of DPs affects the propositional meaning of the sentence:

(1) Filmen   var   *egentligen*   bra.
    Movie-THE is-PAST *PARTICLE*   good.
    'The movie was actually good.'

(2) Filmen   var   *faktiskt*   bra.
    Movie-THE is-PAST *PARTICLE*   good.
    'The movie was actually good.'

In the first sentence, using *egentligen* could imply that for some reason, the person watching the movie did not enjoy it, even though they know the movie should be good. *Egentligen* highlights that there might be another 'truer' truth about the sentence; it can mean that this person found the movie to be bad. In sentence (2), the use of *faktiskt* could imply that the person was a bit surprised by the fact that the movie was good, but it could also be solely confirming that the movie indeed was a good one. *Faktiskt* is used to guarantee the truthiness of a statement, often in light of contradicting prior beliefs ('I was convinced the movie would be bad'), and these prior beliefs are often somehow expressed in the same sentence or adjacently. When no clues to contradictory beliefs are stated, *faktiskt* is often simply confirmatory or augmentative (Teleman et al., 1999). Either way, the person found the movie to be good. The interpretations of these sentences are influenced by previous context, knowledge about the speaker, visual context and other extra-linguistic factors. It tells us something about the speaker's attitude toward the goodness of the movie. At what point does the interpretation of a sentence with a DP in it change drastically enough so that the movie is not considered good anymore? Can a DP impact the propositional content to the point of negation? In the present study, *egentligen* is expected to give rise to an adversative interpretation that negates the content of the sentence (see *eigenlijk* in van Bergen & Bosker, 2018) and will be referred to as an **adversative DP**. In contrast, *faktiskt* should only elicit the confirmatory interpretation, and it will be referred to as a **confirmatory DP** henceforth.

We pose the following research questions:

RQ1: How do the DPs *egentligen* and *faktiskt* affect the interpretation of sentences in low-constraint contexts?

RQ2: How do the DPs *egentligen* and *faktiskt* affect the online processing of sentences in low-constraint contexts?

The experiment was designed so that participants were to listen, look, decide and click on an object. The processing costs induced by DPs were measured online and combined with an offline measure to tap into the processes underlying the interpretation of DPs. This novel design also allowed for comparisons using pupillometry, which, to the best of the authors' knowledge, has not been used in the VWP before.

In the experiment, the adversative (*egentligen*) and confirmatory (*faktiskt*) DPs were compared to a non-ambiguous adverb (*väldigt*). The latter is used as the control condition, as it is a highly frequent adverb, meaning 'very', that does not prompt any ambiguity (see van Bergen & Bosker, 2018). Participants heard a conversation while looking at four objects on a screen. They were told that their task was to figure out which object the conversation was about, and that the speakers would never utter the actual name of the object. Two objects were identical but were either black/white or gray, and two were unrelated distractor objects. Participants heard conversations where the beginning of the conversations would always reveal some sort of clue that made it possible for them to hone in on the two possible objects in question. For example, if the target object was a shirt, they would hear the word *sleeves* before they heard the manipulated line in the conversation. In this line, one of the interlocutors said 'it is *PARTICLE/CONTROL ADVERB* black', and there was a black *target object*, and a gray *competitor object.* When hearing *egentligen*, participants might pragmatically infer the sentence to mean 'it used to be black, but now it has faded and is, in fact, gray'. But, given that the word 'gray' is never uttered in any of the conversations, the most 'logical', semantic interpretation would be to choose the black object (see Bott & Noveck, 2004). The experiment thus tested whether the participants would opt for the pragmatic (gray shirt) or the semantic (black/white shirt) interpretation when *egentligen* is used. We predicted that this adversative particle would slow down reaction times and make the majority of the participants choose the gray shirt over the black shirt. *Faktiskt* should in these conversations only work as confirmatory of the objects' colors, but no prediction about how *faktiskt* would be processed and interpreted was made beforehand.

We opted for an experiment design where the low-constraint contexts reflect a plausible state change, and the participants choose one of two outcomes, depending on how they interpret the DP. While other studies have contrastive elements that represent unlikely or likely outcomes, the choice in this experiment is between two objects that are the same but represent endpoints on two scales of color; black and gray or white and gray. Because the contexts are loosely constraining in the sentences *without* the DP, it is only the presence of a DP that prompts different responses, i.e., choosing the gray competitor object over the black or white target object. We can thereby investigate how DPs affect the processing and interpretation without manipulating contexts to guide the participants down a specific path of interpretation.

3. Methods

3.1. Participants

Forty-two native speakers (33 females) of Swedish were tested. All participants had normal or corrected-to-normal vision and no hearing problems. The participants aged between 19 and 58 years ($M = 28.85$, $SD = 10.19$, $MEDIAN = 25$). All participants were renumerated for participation.
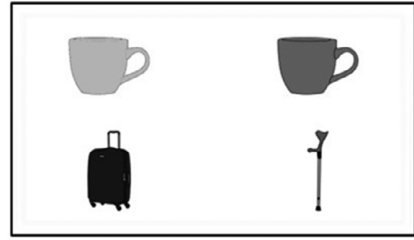
3.2. Design

Experimental materials comprised 36 items with a screen of 4 objects combined with a spoken, everyday conversation (Figure 1). Thirty-six visual displays were
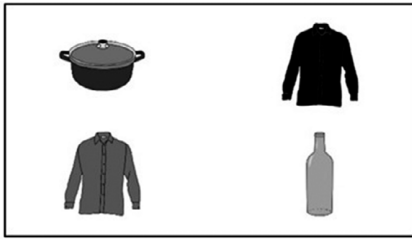
Figure 1. Examples of stimuli. Left: target item, right: filler item.

constructed, each containing four objects that appeared in four corners of a 17.3-inch computer screen. Each display contained two possible referents that were the exact same object that were either black/white and gray, as well as two unrelated referents that were used as distractors. All pairs of target referents (e.g., a black shirt and a gray shirt) were spread evenly across all possible positions on the screen (upper row, lower row, left hand side, right hand side and diagonally going from the upper left corner to the lower right corner, and vice versa). All pictures were selected from the MultiPic database (Duñabeitia et al., 2018), which contains pictures that have been normed for psycholinguistic tasks in 6 different European languages. All pictures were in grayscale but the pictures used for the target referents have been modified to appear white, gray or black depending on the spoken stimuli. (Making alterations to the pictures was approved after communication with the creator of MultiPic.) 15 (approximately 40%) of the target items were manipulated with the colors white and gray, and 21 (approximately 60%) with the colors black and gray. All target objects in critical trials were fabric of some sort, since black fabric quite naturally can fade into gray as it turns older, becoming grayer because of the sun or coloring that fades. White fabric can become visibly rugged or dirty, and therefore appear gray.

The conversations in the target items give the participants enough clues to figure out that the object in question is either the competitor object (the gray shirt) or the target object (the black shirt). It is the DP that introduces a plausible state change in color, suggesting that the color might have faded and turned into gray, or in cases with the white target objects, it somehow has become visibly dirty and therefore gray instead. The conversations average on 5.2 seconds (*MIN* = 2.9 s, *MAX* = 7.6 s, *MEDIAN* = 5.2 s) before the DP, giving the participants plenty of time to identify all four objects on the screen before they have to choose (see Ito & Knoeferle, 2023). After the critical word, the average target conversation had 4.97 words following the critical word (black or white) (*MIN* = 3, *MAX* = 7, *MEDIAN* = 5).

Forty-eight filler items, with only one target object and three distractors, were created following the same structure as the target items. 25% (12) of the filler items contained two distractors of the same shape but different colors. 25% (12) of the filler items had either semantically competing distractors (such as both target and one of

the distractors were vehicles or pieces of furniture e.g.) or visually similar distractors (if the target object was round, then so was one of the distractors). The remaining 50% (24) of the filler items had unrelated distractor objects. As filler items only had one plausible referent, they also worked as a sanity check for participants, and as a control for whether or not they understood the task properly and remained alert throughout the experiment. The clues for the target objects were spread throughout the filler item conversations; sometimes the participants could figure out the target object straight away and sometimes the last word of the conversation was the missing piece of the puzzle, but most often one (or several clues) came somewhere in the middle.

In addition to being fabric of some sort, distractor objects, as well as target objects in filler trials, could also be sports gear, tools, vehicles and other everyday objects, such as calculators, bundles of yarn, coins and so on.

Spoken dialogues were recorded by two female native speakers of the standard variety of Swedish. The speakers were instructed to put a slight emphasis on the color words (*black* or *white*) that succeeded the critical words (*egentligen*, *faktiskt* or *väldigt*) to ensure that no emphasis was put on the discourse particles or the control adverb. All conversations were recorded twice so that both speakers read every part. The best recordings were then chosen. The conversations were segmented using Audacity (Audacity Team, 2022).

Experimental items were counterbalanced across three lists using a Latin square design, such that each participant encountered each conversation in only one condition (12 items per condition). The 48 filler items were added to each list, yielding a total of 84 conversations distributed over 3 blocks, where one block contained a total of 28 items – 12 target items and 16 filler items. The trial order was pseudorandomized such that the participant saw a maximum of two target items in sequence but never two target items of the same condition after each other. Each participant saw the blocks in one out of three possible orders.

During the first phase of each trial (Figure 2), gaze and pupillometry data were collected. During the fixation cross, pupillometry data were collected. During the third phase, reaction times and responses were collected. The experiment started with three screens of written instructions, where the participant clicked on the mouse to
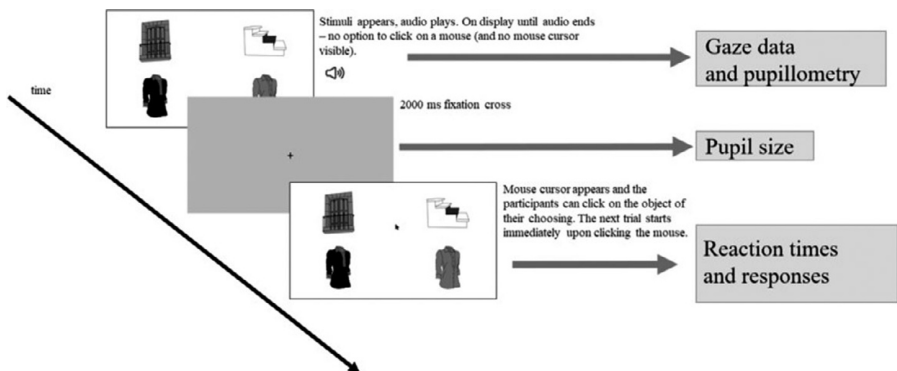

Figure 2. Experiment design: trial work flow. For each trial, the screen with four objects was displayed while the audio unfolded. Once the audio had played out, a 2000 ms gray screen with a fixation cross was presented. Then the four objects appeared again, and participants were allowed to select an object using the mouse.

continue to the next screen. After the instructions, a 9-point calibration was run. Participants then completed three practice trials and were able to ask questions before the experiment started. The experiment proceeded with three blocks divided by two self-paced pauses.

3.3. Procedure

Participants were tested at the Uppsala Child and Baby lab at Uppsala University. All participants gave written consent to participate in the experiment. Prior to starting the experiment, participants filled out a background questionnaire. The eye-tracker Tobii pro Nano was used, and the experiment took approximately 20 minutes to conduct.

The study was reviewed by the Swedish Ethical Review Authority (dnr 2023–00963-01). The authority's recommendation was that no ethical vetting was necessary for the project. The study was pre-registered at aspredicted.org on May 31, 2023 (dnr 134030). This research was funded by The Royal Gustavus Adolphus Academy for Swedish Folk Culture and the IDO foundation for language research in memory of Hellmut Röhnisch.

3.4. Analysis

Statistical differences in behavioral responses and reaction times were assessed using logistic and linear mixed-model regression, respectively (Figure 3), using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2023). Models included binary responses (Target or Competitor) or log transformed reaction times as dependent variables and Condition (Adversative, Confirmatory, Control) as the independent variable. Stimulus list was included as a covariate. Random intercept for participant and trial was included in the model. Statistical significance was computed using stepwise model comparisons (using the *lmerTest* package in R; Kuznetsova et al., 2017) for reaction times, and pairwise comparisons and marginal means were estimated using the *emmeans* package in R (Lenth, 2024).

We used cluster-based permutation statistics to identify points in time where the participants diverged in their proportion of looks to Targets and Competitors (Figures 4 and 5). Using this method, clusters of adjacent time bins were compared to a permutation distribution of the maximum cluster statistics from repeated random sampling (N = 1000), from which statistical significance was derived. This non-parametric test of statistical significance controls for both multiple comparisons and autocorrelation between measurements (Maris & Oostenveld, 2007). The dependent variable was the log ratio proportion of looks to Target and Competitor objects (Ito & Knoeferle, 2023).

Pupil size is a reliable indicator of cognitive load, which is the mental effort used in working memory. This connection is due to the autonomic nervous system reacting to mental demands. Research dating back to 1966 by Kahneman and Beatty has shown that pupils dilate with increased cognitive effort, a finding consistently supported by later studies (van der Wel & van Steenbergen, 2018). Additionally, pupil dilation is linked to activity in the Locus Coeruleus, which handles noradrenaline production affecting attention and arousal. This makes pupil size a useful tool for assessing cognitive load in real-time, across the lifespan (Laeng, et al., 2012).
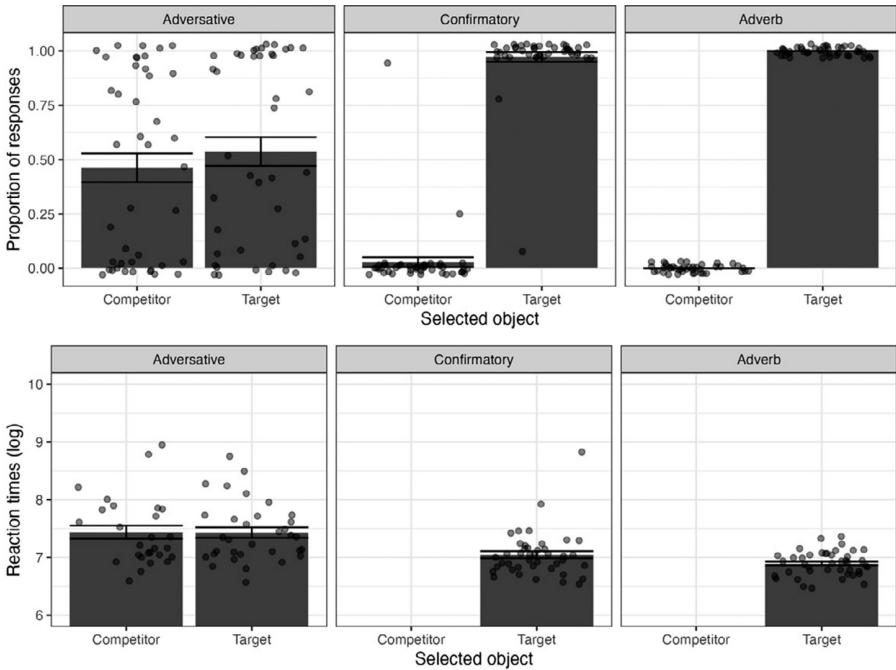
Figure 3. Upper panel shows Target and Competitor responses for each condition. Each dot represents the mean of participants, where 1 corresponds to the participant always choosing the Target object, and 0 corresponds to always choosing the Competitor object. The lower panel shows reaction times and individual mean response latencies. Error bars show standard error of the mean.

In order to assess cognitive load between the different conditions, we measured pupil size following the onset of the condition word (DP/control adverb) (Figure 6). Raw pupil measurements were smoothed using a moving Hanning window (N = 5). To control for differences in offset between participants, individual responses were baseline corrected by subtracting the mean pupil size during the 500 ms period immediately preceding the onset of the discourse particle/control adverb from all data points in each trial. Preprocessing was conducted in Tobii Pro Lab (Tobii, 2021).

Because pupillary responses may be susceptible to differences in light contrast in the Target and Competitor objects, we also measured pupil size during the fixation cross period of 2000 ms and compared this to the mean pupil size in the 500 ms preceding the critical word (*egentligen*, *faktiskt* or *väldigt*) (Figure 7).

All materials, data and analysis scripts are available through Open Science Foundation, available at the following URL: https://osf.io/fdnpw/

4. Results

4.1. Responses and reaction times

Participants tended to select the Target object (83.7% of all critical trials, N = 1512), but the response pattern differed between the conditions. Participants typically selected the Target object in the Confirmatory (97.2%, N = 504) and Adverb (100%, N = 504) conditions, but they were equally likely to select the Target
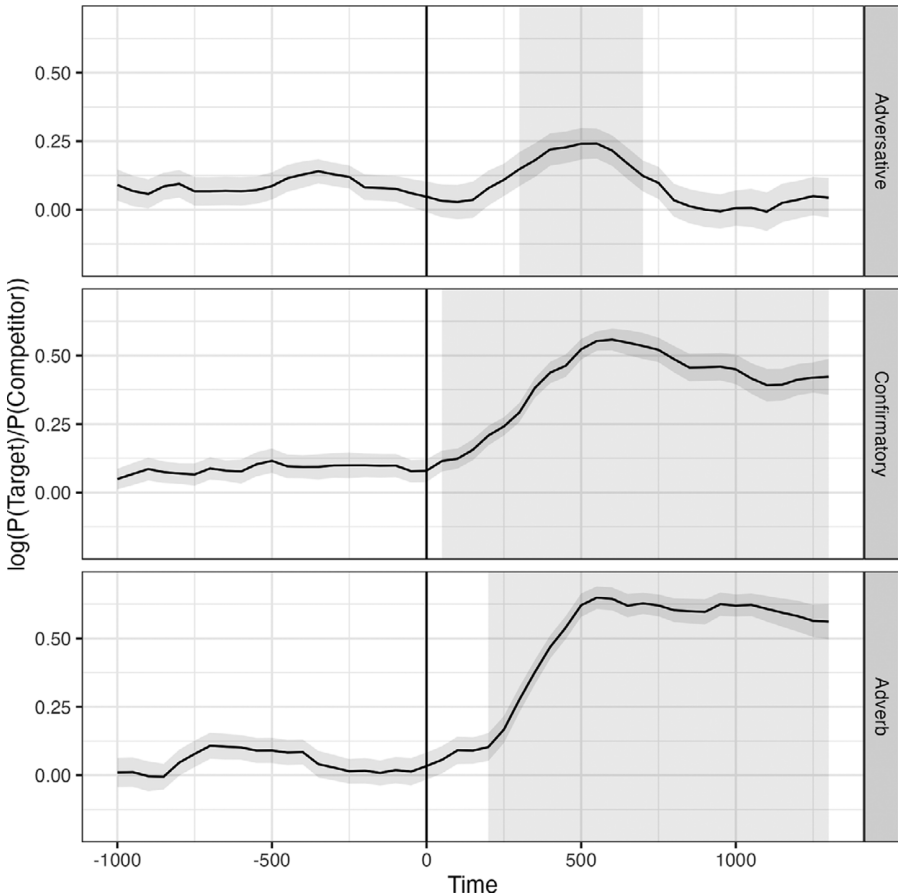
Figure 4. Log-likelihood of looking toward the Target (black shirt) or the Competitor (gray shirt), in the three different conditions: Adversative (*egentligen*), Confirmatory (*faktiskt*) and Adverb (*väldigt*). 0 on the X-axis marks when the participants heard the word 'black' or 'white'. Positive values indicate more looks toward the Target, and negative values indicate more looks toward the Competitor.

(53.8%, N = 504) and Competitor (46%, N = 504) objects in the Adversative condition ($t(57.86) = 0.363$, $p = 0.71$; Figure 3, upper panel). In the Adversative condition, some participants always selected either the Target object (N = 17) or the Competitor object (N = 13), with over 90% probability of selecting one of the objects across all trials. Some participants (N = 12), however, remained undecisive and selected Target and Competitor objects on different trials.

Responses in Target trials were slower (Figure 3, lower panel) in the Adversative compared to the Confirmatory condition ($\beta = 0.214$, $SE = 0.043$, $t_{ratio} = 4.993$, $p < .001$) and the Adverb condition ($\beta = 0.325$, $SE = 0.043$, $t_{ratio} = 7.605$, $p < .001$), but latencies did not differ between the Confirmatory and the Adverb conditions ($\beta = 0.111$, $SE = 0.05$, $t_{ratio} = 2.233$, $p = 0.07$). There was no difference in response latency to Competitor and Target objects in the Adversative condition ($\beta = -0.092$, $SE = 0.074$, $t_{ratio} = -1.252$, $p = 0.21$).
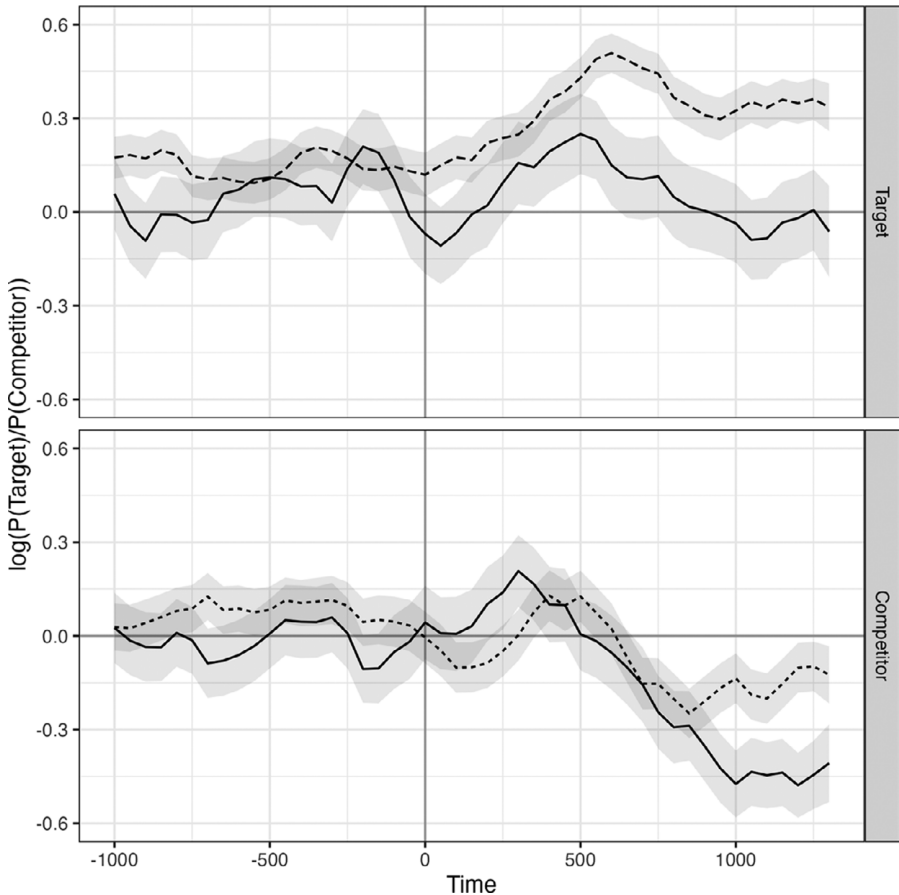
Figure 5. Click-contingent gaze patterns in the Adversative condition when selecting the Target object (Upper panel) or the Competitor object (Lower panel), by participants who consistently selected either the Target object (dashed line) or Competitor object (dotted line), or remained undecisive between the two (solid line). Ribbons show standard error of the mean.

4.2. Gaze patterns

The behavioral responses were reflected in gaze patterns (Figure 4), where partici-pants looked toward the Target object in the Confirmatory and Adverb conditions within 500 ms after the critical word, while they remained undecisive for the duration of the trial in the Adversative condition. Click-contingent responses in the Adver-sative condition (Figure 5) show that participants who consistently selected the Target or Competitor objects looked at the selected object early in the trial, while the ambiguous responder group only did so when they selected the Competitor object, but they remained undecisive longer when they selected the Target object.

4.3. Pupillometry

Pupillary responses (Figure 6) show that the pupil dilated and remained significantly larger than baseline throughout the Adversative trials, while in the Confirmatory
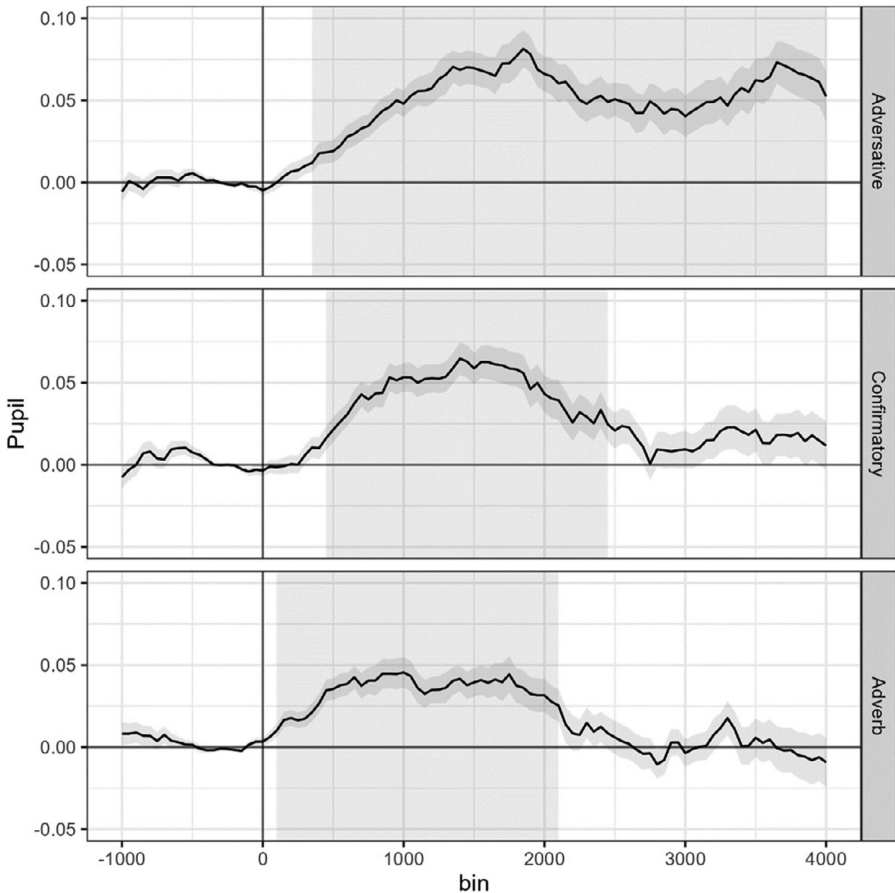
Figure 6. Baseline corrected pupillary responses. Gray boxes indicate sections in time where pupillary responses differ from baseline using cluster-based permutation statistics. Ribbon shows standard error of the mean. 0 on the X-axis corresponds to the point in time when they heard the DP or the adverb.

trials, pupil dilation was sustained for approximately 2000 ms before it returned to baseline. Because pupillary responses may be susceptible to differences in brightness and contrast in the Target and Competitor objects, we also measured pupil size during the fixation cross period of 2000 ms and compared this to the mean pupil size in the 500 ms preceding the discourse particle/adverb (Figure 7).

Only in the Adversative condition did the response differ from baseline ($M = 0.043$, $CI = [0.016\ 0.070]$) while in the Confirmatory ($M = 0.011$, $CI = [-0.016\ 0.039]$) and Adverb ($M = -0.002$, $CI = [-0.03\ 0.026]$) conditions, it did not. There was also a significant difference between the Adversative and Adverb conditions ($\beta = 0.05$, $SE = 0.017$, $t_{ratio} = 2.713$, $p = 0.02$), but not between the other conditions (Adversative versus Confirmatory: $\beta = 0.03$, $SE = 0.017$, $t_{ratio} = 1.1927$, $p = 0.137$; Confirmatory versus Adverb: $\beta = 0.01$, $SE = 0.017$, $t_{ratio} = 0.787$, $p = 0.712$). There was no difference in pupil dilation between participants who selected the Target or Competitor objects, and those who alternated between both during Adversative trials ($ps > 0.21$). These results
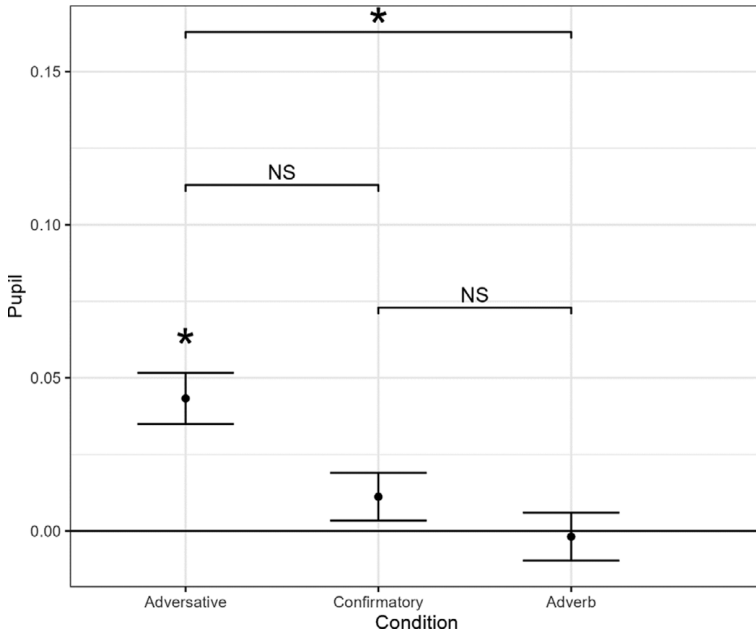
Figure 7. Pupillary dilation during the fixation period averaged over participants. The star indicates where pupil dilation differs significantly from baseline. Error bars show standard error of the mean.

corroborate the pattern in Figure 6: the participants' pupil dilation is the largest in the Adversative condition, while neither the Confirmatory nor the Adverb conditions differed from baseline.

5. General discussion

Discourse particles are crucial in successful communication and are frequently used to help express what is not explicitly stated. In previous studies, the processing and interpretation of DPs have been studied by constraining the contexts in order to specifically invoke the canonical interpretations that the DPs are assumed to give rise to, but results have been inconclusive. In this study, we instead investigate how the introduction of DPs into low-constraint contexts affects interpretation and processing. We examined how Swedish adversative (*egentligen*) and confirmatory (*faktiskt*) DPs affect the interpretation and the online processing of sentences as they unfold, as compared to the control adverb *väldigt* ('very').

5.1. Offline and online measures reveal distinct response strategies

Pertaining to the interpretation of sentences in low-constraint contexts (RQ1), and in line with our predictions, participants always selected the (black or white) Target object in the Confirmatory (*faktiskt*) and Adverb (*väldigt*) conditions. For *egentligen*, we predicted that participants would choose the gray Competitor object on a majority of trials, but this was not the case. Instead, participants used three different strategies: two thirds of the participants consistently chose either Target or

Competitor, while the remaining participants alternated between the two options throughout the experiment.

How these particles affect the online processing (RQ2), is visible in results from the reaction times, gaze patterns and pupillometry. For *faktiskt* and *väldigt*, there was no difference in response latency. Since response patterns and response latencies are similar, this suggests that *faktiskt* did not induce any processing facilitation as compared to *väldigt*. Gaze patterns provide deeper insight into the three different response strategies in the Adversative condition with the DP *egentligen*. One explanation is that participants opted for the semantic interpretation (see Bott & Noveck, 2004; Huang & Snedecker, 2009), and chose the black or white object because *black* or *white* are words that were explicitly mentioned in the conversations. There are two plausible pragmatic interpretations of the DP *egentligen*, in which the listener might consider the gray object even though the word is never mentioned in the conversations. One is the interpretation 'in this light it looks gray, but it is *in fact* black/white', which would make participants choose the Target object, even though they look at the plausible gray object at first. The other pragmatic interpretation, which is in line with our predictions, is that the participants chose the Competitor object because they interpreted *egentligen* to have the meaning 'this object used to be black/white, but has now turned gray'.

Click-contingent gaze patterns indicate that the only group of participants who quickly stopped looking at the gray Competitor object were the ones who consistently chose the black or white Target object (N = 17). Since this group stop looking at the Competitor object all together, we conclude that they opt for the semantic interpretation; they hear the word black/white, look at the black/white object and end up choosing this object. Participants who consistently chose the gray Competitor object (N = 13) looked at the Target object at first but quickly fixed their gaze on the Competitor object. This indicates a pragmatic interpretation; the object used to be black/white but has now turned gray. The ambivalent group (N = 12), participants who alternated between Target and Competitor, continuously looked at the Competitor object even in the cases where they ended up choosing the Target object. We suggest that they might shift between the two different pragmatic interpretations that both include the possibility of the object being gray.

These two latter groups, a majority of the participants (N = 25), consistently entertained the option of the gray Competitor object regardless of their actual response. This means that the gaze pattern fully supports our predictions that the participants would consider the gray object a majority of the time, which indicates that *egentligen* has an adversative function. This adversative function would not be detectable had we only looked at reaction times and responses, and speaks to one of the strengths of using eye tracking, where multiple online measures that tap into processing can be collected with ease.

Egentligen also induces slower reaction times, and all participants show a significant pupil dilation and this effect was sustained throughout the trial and during the fixation cross. In the other conditions, pupil size quickly returned to baseline and did not differ from baseline during the fixation cross. Upon hearing *egentligen*, participants struggle with what to choose. This is shown in the reaction times, in the enlarged pupils, in the gaze patterns and in the responses. Even the participants who opt for the semantic option and always choose black or white are slower and have enlarged pupils. For the Confirmatory condition, the reaction times, pupil size, gaze patterns, as well as responses, are the same as for the control

Adverb. Had there been a facilitatory effect of *faktiskt*, it should have shown in reaction times and/or pupil size. This unequivocally shows that *egentligen* brings about a higher cognitive load, while *faktiskt* does not show any signs of facilitation. If anything, there is a slight pupil dilation visible in the Confirmatory condition as compared to the control condition, albeit not significant. This numerical trend suggests participants find adversative DPs the hardest to process, then the confirmatory and lastly, the unambiguous control adverbs.

Even though the confirmatory *faktiskt* has similar results as the control adverb *väldigt*, the lack of facilitation in the Confirmatory condition in itself can be informative. We suggest that this lack of facilitation related to the confirmatory DP *faktiskt* arises because the integration of DPs into the processing requires an immediate reanalysis of made assumptions. Before the participants hear the DP and the color, they have already figured out that their choice is between two objects; either black/white or gray. No matter if the DP confirms the color word that is uttered, participants still have to reanalyze their assumption because *faktiskt* can mean so many different things, along the lines of 'actually' and 'really'. It can often convey a surprisal toward the propositional content, similar to how 'actually' can be interpreted. The participants are left to wonder why the speaker would say *faktiskt*, trying to 'fill in the blanks' from what is said before in the conversation. This is in line with the claims of Filik et al., (2009), that DPs that convey surprising facts are more complex and therefore more strenuous to process. However, the reanalysis of *faktiskt* does not lead to an interpretation shift, as being surprised by the color of an object would not change the color, and *faktiskt* can also simply mean 'really'.

5.2. DPs increase under-specificity in low-constraint contexts

The results limit the findings in Loureda et al. (2022), who claim that DPs always facilitate processing. DPs that do this to the largest degree are counter-argumentative, e.g., *sin embargo* ('however') and *a pesar de ello* ('despite that'). The adversative qualities of *egentligen* are similar to the functions of counter-argumentative DPs. So why then, are the results from the present study diametrically different? We suggest that this discrepancy stems from how contexts are construed and manipulated in the different experimental stimuli. The reading paradigm experiment in Loureda et al. (2022) was designed with utterances that either contained DPs or not, but the present study investigated DPs in comparison to an unambiguous adverb in a Visual World paradigm. The stimuli in Loureda et al., (2022) consisted of sentences that without the DPs were analogously linked, and the participant had to pragmatically infer what the DPs usually help to infer. See the example below:

> SP. Enrique y Marta estudian mucho. **Sin embargo/A pesar de ello**, sacan malas notas. *'Enrique and Marta study hard. **However/Despite that**, they get bad grades.'* (Loureda et. al., 2022, p. 24)

This utterance is more difficult to process without the DP. It becomes *under-specific* if the DP is not there and the reader has to pragmatically infer the relationship between studying hard and getting bad grades. The example should be easier to process with the counter-argumentative DP in it. This might be the reason why Loureda et al. (2022) found the largest effect for the counter-argumentative DPs. The contexts without the counter-argumentative DPs were the hardest to process because it is the

biggest leap the hearer needs to make on their own, and one that changes the state of affairs drastically. Similarly, *only* and *even* are contrasted with unlikely or likely continuations, and differ in what processing costs they exert depending on if the contexts support their most common interpretation (*even* with *unlikely* and *only* with *likely*) or not (Filik et al., 2009).

In the present study, the under-specificity increases when the DP is added; the difference between *den är väldigt svart* ('it is very black') and *den är egentligen/faktiskt svart* ('it is *DP* black') is that the first sentence gives enough information for it to be easily understood and the task to be resolved; the participant simply chooses the black shirt, as *väldigt* only strengthens the color word. The other sentences give rise to processing difficulties. Due to its procedurality and reflexivity, a DP is too strong of a cue to be ignored. But there are not enough cues in the surrounding context for the DP to have a facilitatory effect on decoding the meaning. In accordance with Kissine and De Brabanter's (2023) claim about the classic Gricean relevance theoretic informativeness, this could be a case of the participants being held up by trying to figure out what the speaker means by saying *egentligen* or *faktiskt*. Speakers would not say the DP, if it was clear that the shirt is black. When the context is not maximally relevant, DPs therefore increase processing costs instead of facilitating the process.

Investigating the predictive function of DPs, van Bergen and Bosker (2018) and Rasenberg et al., (2020) assume that the Dutch cognate of *egentligen*, *eigenlijk*, has a core meaning of being adversative and should signal upcoming unexpected information. They found immediate integration but no facilitatory effect. Perhaps the introduction of the DP did not sufficiently restrict the interpretation to a solely adversative one, rendering it under-specified and more difficult to process. In van Bergen and Bosker, (2018) there was great variability and a slow-down upon encountering *eigenlijk*, similar to the results of the present study. The core meaning sought for in previous studies might thus be a mere reflection of the most common interpretation in high-constraint contexts, more in line with generalized conversational implicatures (GCI), first introduced by Levinson (2000). GCI:s are pragmatically inferred interpretations that have become 'preferred' meanings due to frequent occurrence. Both *eigenlijk* and *egentligen* often prompt an interpretational route that leads to an adversative conclusion but not all the time.

5.3. The obligatory interpretation of DPs triggers immediate reanalysis

As one felicitously used DP gives maximum cognitive effect (giving several sentences worth of information) for minimum cognitive effort (saying the DP) (Loureda et al., 2022), the opposite seems to be true for an *infelicitously* used DP. The procedural quality of DPs means that they are so packed with information that even when they give too little information, participants need to start thinking about what they can mean, causing a large cognitive effort. This firmly supports previously stated characteristics, such as DPs serving fundamental communicative needs (Hogeweg et al., 2016) and being communicatively obligatory (Diewald, 2010). We suggest that the increased processing costs visible in the online data is evidence of an increased cognitive load that is driven by the obligatory interpretation of these particles. Older beliefs in the field, where DPs were considered 'extra-linguistic' and redundant (see Alami, 2015) can therefore be further refuted.

The results of the present study both support and contradict the cognitive principles postulated in Loureda et al. (2022). DPs do modify the processing strategy, as they cannot be ignored. Whether they set a maximum of processing costs and immediately regulate the interpretations of the segments they affect, seem to be a matter of context. In the cases described and tested in Loureda et al. (2022) they do, but in our study, they introduce ambiguity that increases processing costs compared to when unambiguous adverbs are used. DPs do not offer optimal control of the reanalysis of utterances that becomes necessary in their presence, and again, this seems to be context-dependent. The Adversative condition in our study opens up for several different interpretations of the segment it affects. The results are in line with relevance theoretic claims (Sperber & Wilson, 2002) about pragmatic inference and processing costs.

While Loureda et al. (2022) investigated the impact DPs have in optimal contextual conditions, the present study investigated how DPs affect processing in contexts where the presence of the DPs does not narrow the meaning. The diverging results could be explained through the inferential account of lexical adjustment, proposed by Wilson and Carston (2007). In search of the most relevant interpretation of a sentence, the hearer will use inferences to either broaden or narrow the possible interpretations, creating an *ad hoc concept* for each new ambiguous utterance they hear. The hearer's goal is to arrive at the proper interpretation as quickly as possible at the lowest possible cost. Only if the initial assumption does not meet the criterion of being the most plausible interpretation does the hearer reanalyze it. In the case of the confirmatory *faktiskt* and the unambiguous *väldigt*, the hearer can stand by the initial assumption of a black shirt, whereas the adversative sentiment of *egentligen* is too strong to be ignored, and could mean that the shirt has become gray. Wilson and Carston (2007) point out that ad hoc concepts encompass both narrowing and broadening, and indeed contribute to the truth-conditional content of utterances, which is the case with *egentligen* when it contributes to the negation of the color of the shirt (it has become gray instead). This unifying account could explain why DPs in certain contexts increase processing costs, while they facilitate processing in other, more specified contexts.

Pragmatic inference does not automatically induce higher processing costs, because such an interpretation can be computed effortlessly when the context provides enough evidence for the interpretation to be plausible. If the context does not provide enough support for a pragmatic interpretation that could be relevant, however, extra processing costs will occur. This might be the simple explanation as to why results about the processing costs induced by DPs diverge. Processing costs decrease when contexts allow for a clear-cut interpretation of the sentence with a DP in it (see Loureda et al., 2022), but they increase when a DP gives rise to several plausible interpretations by introducing additional information that needs to be processed (see Canestrelli et al., 2013; Filik et al., 2009) or if they increase ambiguity, as is the case in the present study.

We propose that the immediate integration supported in previous studies in fact reflects the immediate *reanalysis* that a DP prompts. When the reanalysis is quicker than the analysis of the entire sentence would be without the DP, then this looks like processing facilitation in online data. When it is slower than it would otherwise be to process the sentence without the DP, this looks like processing disadvantages. This reanalysis involves three dimensions that need to be processed: i) the linguistic intuition about the DP; ii) the assumptions made about speaker intention and iii) contextual surrounding. The first dimension involves the semantic interpretation of

the DP; what does this word (most often) mean to the hearer? How does the hearer use it itself most frequently? In other words, what are their GCI (Levinson, 2000) of the word? The second dimension pertains to the maxim of quantity (Grice, 1989) and the communicative principle (Sperber & Wilson, 2002). The hearer needs to take speaker intention into account here. What does the person mean by saying the DP? And finally, the third dimension – contextual surrounding. What other contextual cues are there to support the interpretation of the DP? What has probably been said before?

There is evidence of these three dimensions being taken into account in the present study. The linguistic intuition about *egentligen* prompts a majority of the participants to look at the gray Competitor object, whereas the presence of *faktiskt* does not induce looks to the Competitor object, confirming that the most common interpretation of *egentligen* is adversative, and that *faktiskt* is confirmatory. The pupil dilation visible for both *egentligen* and *faktiskt* speaks to the immediate reanalysis made upon hearing a DP, and the difference in how much the DP affects the propositional content (*egentligen* could possibly negate it while *faktiskt* confirms it) explains the differences in magnitude of these online measures, and is in line with previous results on surprising information/drastic change/unexpectedness (Canestrelli et al., 2013; Filik et al., 2009; van Bergen & Bosker, 2018). This also speaks to the fact that the participants have to consider why a speaker would say the DP. If not, the confirmatory *faktiskt* should work facilitatory and speed up processing as compared to the adverb. Finally, we believe that the differences between the online measures, where a majority of the participants entertain the pragmatic interpretation in the Adversative condition, and the offline measures of their responses, where the gray object was chosen only half of the time, reflects the third dimension. The participants who chose the black object all the time quickly decided that the literal semantic meaning of *black* trumps pragmatically driven interpretations. The ambivalent group who alternated between the gray objects and the black objects probably weighed between the one meaning *egentligen* can have, 'in fact', and the adversative sentiment it can convey. In order to resolve this ambiguity, other contextual factors have come to play (e.g., is the object old or new?). The reflexive quality DPs have (Aijmer & Simon-Vandenbergen, 2004) also plays into this third dimension; hearers are aware that speakers use DPs to refer to local context and perhaps alter the context, therefore they need to pay extra attention to the details in the context. One cannot automatize all interpretations a DP gives rise to. Therefore, all plausible interpretations that can be derived from a given context needs to be considered in each new example, taking all semantic cues in the particular item into account (see Canestrelli et al., 2013). We suggest that integrating information about these three dimensions is simultaneous and rapid. The slower reaction times for all participants, no matter which object they ended up choosing in the Adversative condition, also suggest that resolving the reanalysis for *egentligen* was the most strenuous and time-consuming one. The variability among responses and gaze patterns combined with the longer reaction times, suggest that some participants' linguistic intuition about particular words weighs heavier than context and vice versa.

6. Summary

The combination of pupillometry, gaze patterns, reaction times and offline measures in the present study enabled a fine-grained analysis of the processing and a detailed

picture of how DPs affect interpretation and processing. The fact that a majority of the participants entertain the possibility of the object being gray in the Adversative condition speaks to the multifunctionality of *egentligen.* It introduces the possibility of a propositional change (the shirt is *not* black), even though the word *gray* is never mentioned throughout the experiment. Blakemore (2002) points out that the distinction between truth and non-truth conditionals is not relevant anymore. But it is of interest that DPs can impact a sentence to a degree where it negates the propositional content. This should be further investigated, perhaps adjacently to linguistic forms that indirectly work to negate propositional content. In future lines of inquiry, the dimension of under-specificity versus too highly constrained contexts should also be controlled for in the design of the experimental stimuli. Using a DP is a powerful tool to help guide the interpretation process, and this study finds that immediate integration (see Grodner et al., 2010; Kurumada et al., 2014; Loureda et al., 2022), is in fact immediate *reanalysis.* This reanalysis is prompted because of the procedural and reflexive qualities of DPs and depends on the contexts and the increased under-specificity that the DPs in this experiment induce. DPs cannot be ignored even though they do not help, and participants are left to wonder why the speaker would say *egentligen* or *faktiskt*, which is connected to the maxim of quantity (Grice, 1975/1989). Furthermore, additional processing costs seem to be linked to the complexity of the reanalysis; as is the case with *eigenlijk* and processing delays (van Bergen & Bosker, 2018), *sogar* and drastic change (Gerwien & Rudka, 2019), the subjective *want* and the need for perspective change on behalf of the hearer (Canestrelli et al., 2013), and *even* with information that is unexpected or surprising (Filik et al., 2009). We propose that three dimensions are involved in the reanalysis, all triggered at the onset of the DP and simultaneously activated: i) the participant's linguistic intuition about the meaning of the word, ii) the assumptions made about speaker intention and iii) the impact of the contextual surrounding on the utterance.

References

Aijmer, K., & Simon-Vandenbergen, A. M. (2004). A model and a methodology for the study of pragmatic markers: The semantic field of expectation. *Journal of Pragmatics*, 36(10), 1781–1805. https://doi.org/10.1016/j.pragma.2004.05.005

Alami, M. (2015). Pragmatic functions of discourse markers: A review of related literature. *International Journal on Studies in English Language and Literature*, 3(3), 1–10.

Audacity Team. (2022). *Audacity (Version 3.2.2) [Computer software]*. https://www.audacityteam.org/

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

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486456

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 51(3), 437–457. https://doi.org/10.1016/j.jml.2004.05.006

Canestrelli, A. R., Mak, W. M., & Sanders, T. J. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes*, 28(9), 1394–1413. https://doi.org/10.1080/01690965.2012.685885

Carston, R. (1999). The semantics/pragmatics distinction: A view from relevance theory. In K. Turner (Ed.), *The semantics/pragmatics interface from different points of view* (pp. 85–125). Elsevier Science.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: a constraint-based approach. *Cognitive Science*, 39(4), 667–710. https://doi.org/10.1111/cogs.12171

Diewald, G. (2010). On some problem areas in grammaticalization studies. In K. Stathi, E. Gehweiler, & E. König (Eds.), *Grammaticalization: Current views and issues*, (Studies in Language Companion Series, 119, pp. 17–50). John Benjamins. https://doi.org/10.1075/slcs.119

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, 71(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Filik, R., Paterson, K. B., & Liversedge, S. P. (2009). The influence of only and even on online semantic interpretation. *Psychonomic Bulletin & Review*, 16, 678–683. https://doi.org/10.3758/PBR.16.4.678

Gerwien, J. & Rudka, M. (2019). Chapter 8. Expectation changes over time: How long it takes to process focus imposed by German sogar. In Ó. Loureda, I. Recio Fernández, L. Nadal, & A. Cruz (Eds.), *Empirical studies of the construction of discourse* (pp. 229–252). John Benjamins Publishing Company. https://doi.org/10.1075/pbns.305.08ger

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Volume 3. Speech acts* (pp. 41–58). Academic Press.

Grice, P. (1989). *Studies in the way of words*. Harvard University Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, 116(1), 42–55. https://doi.org/10.1016/j.cognition.2010.03.014

Hogeweg, L., de Hoop, H., Ramachers, S., van der Slik, F., & Wottrich, V. (2016). The L2-acquisition of the German particle doch. *International Review of Applied Linguistics in Language Teaching*, 54(3), 201–227. https://doi.org/10.1515/iral-2016-9981

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, 58(3), 376–415. https://doi.org/10.1016/j.cogpsych.2008.09.001

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, 55(7), 3461–3493. https://doi.org/10.3758/s13428-022-01969-3

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Kissine, M., & De Brabanter, P. (2023). Pragmatic responses to under-informative some-statements are not scalar implicatures. *Cognition*, 237, 105463. https://doi.org/10.1016/j.cognition.2023.105463

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, 133(2), 335–342. https://doi.org/10.1016/j.cognition.2014.05.017

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, 7(1), 18–27. https://doi.org/10.1177/1745691611427305

Lenth, R. V. (2024). *emmeans: Estimated marginal means, aka least-squares means (R package version 1.10.0) [Computer software]*. Available at https://CRAN.R-project.org/package=emmeans.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press. https://doi.org/10.7551/mitpress/5526.001.0001

Loureda, Ó., Fernández, I. R., Cruz, A., & Rudka, M. (2022). Principles of discourse marking: An experimental approach of general and contrastive perspectives. In Ó. Loureda *et al.* (Eds.), *Discourse markers in interaction: from production to comprehension* (pp. 17–44). De Gruyter. https://doi.org/10.1515/9783110790351-002

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Parola, A., & Bosco, F. M. (2022). An eye-tracking investigation of the cognitive processes involved in the comprehension of simple and complex communicative acts. *Quarterly Journal of Experimental Psychology*, 75(10), 1976–1995. https://doi.org/10.1177/1747021822107962

R Core Team. (2023). *R: A language and environment for statistical computing [Computer software]*. R Foundation for Statistical Computing. https://www.R-project.org/

Rasenberg, M., Rommers, J., & van Bergen, G. (2020). Anticipating predictability: an ERP investigation of expectation-managing discourse markers in dialogue comprehension. *Language, Cognition & Neuroscience*, 35(1), 1–16. https://doi.org/10.1080/23273798.2019.1624789

Silverstein, M. (1976). Shifters, linguistic categories, and cultural description. In K. Basso & H. Selby (Eds.), *Meaning in anthropology* (pp. 11–55). University of New Mexico Press.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Blackwell.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, 17(1–2), 3–23. https://doi.org/10.1111/1468-0017.00186

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, 268, 1632–1634. https://doi.org/10.1126/science.7777863

Teleman, U., Hellberg, S., & Andersson, E. (1999). *Svenska Akademiens grammatik* (Vols. 1–4). Norstedts Ordbok.

Tobii. (2021). *Tobii Pro Lab (Version 1.171) [Computer software]*. Tobii AB. https://www.tobiipro.com/product-listing/tobii-pro-lab/

van Bergen, G., & Bosker, H. R. (2018). Linguistic expectation management in online discourse processing: An investigation of Dutch inderdaad 'indeed' and eigenlijk 'actually'. *Journal of Memory and Language*, 103, 191–209. https://doi.org/10.1016/j.jml.2018.08.004

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, 25(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Wilson, D., & Carston, R. (2007). A unitary approach to lexical pragmatics: relevance, inference, and ad hoc concepts. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 230–260). Palgrave Macmillan.