Discourse-level implicature: A case for QUD*

Katja Jasinskaja   Fabienne Salfner   Constantin Freitag

Abstract

This paper argues that multi-sentence discourses give rise to Gricean quantity implicatures that go beyond the mere sum of the implicatures of the sentences they consist of. We formulate two theories of discourse-level implicature: the null theory, which only has a mechanism for sentence-level implicature and does not rely on any specific notion of discourse structure; and a theory that assumes that discourse is hierarchically structured by questions under discussion (QUD) and that QUDs can guide the derivation of quantity implicatures at all levels of discourse structure. In two experiments using the inference task paradigm and focusing on sequences of sentences with contrastive topic, the QUD-based theory is shown to make more accurate predictions than the null theory. This finding provides additional motivation for the QUD-based approach to discourse structure.

1 Introduction

The goal of this paper is to show how the notion of Question under Discussion (QUD) can be employed to describe the phenomenon of *discourse-level implicatures*.

In the classical picture of pragmatics going back to Paul Grice (1975), implicatures are defeasible inferences that result from reasoning about what the speaker has said, has not said, could have said, and what she must have meant, on the assumption that she behaves cooperatively and observes the Maxims of Conversation (Quality, Quantity, Manner). The input to this reasoning process is *the speech act*, and the usual tacit assumption is that a speech act has roughly the size of one sentence. As a result, the bulk of work on implicatures talks about implicatures of individual sentences, whereas implicatures of multi-sentence discourses have been almost entirely ignored in pragmatics research so far. However, it is a common view in discourse semantics that a discourse is itself a complex speech act that consists of simple speech acts (see e.g. Asher and Lascarides, 2003). On this view, one should expect Gricean reasoning to apply to single-sentence speech acts and multi-sentence speech acts alike, so discourse-level implicatures should be generated in much the same way as sentence-level implicatures.

Geurts (2007, 2010) has recently advocated this position using examples like (1). The one-sentence answer (1-a) and the two-sentence answer (1-b) give rise to the same *exhaustivity* inference: Cleo visited Naples, Rome, and Ravenna, *and no other places*. Geurts argues that in (1-b) this must be the result of Gricean reasoning with multi-sentence speech acts.

(1)   Tony: Which places did you see on your trip to Italy?
       Cleo: a.   I went to Naples, Rome, and Ravenna.
                b.   Julius and I first went to Naples and Rome together.
                      Then, while he went to see Milan, I visited Ravenna.

This paper addresses the question of how discourse-level implicatures like that in (1-b) come about. Do we really need to apply Gricean reasoning to multi-sentence speech acts in order to account for those pragmatic effects? Concentrating on exhaustivity, in Section 2 we will consider two simple theories: (a) a *null theory*, which only has a mechanism for sentence-level implicatures and where discourse-level implicatures can only result from combinations of sentence-level implicatures; and (b) a theory of the kind envisaged by Geurts (2010), where implicatures can be generated at all levels of discourse structure. As will become clear, the QUD-based notion of discourse structure provides the most straightforward basis for an extension of the notion of implicature to discourse level in a theory of the latter kind.

On the one hand, we will argue that the null theory, in fact, goes a long way, and a broad class of apparent discourse-level implicatures, including Geurts' example (1-b), can be accounted for without any reference to Gricean reasoning at discourse level. If the null theory were true, there would be no need for QUD-based discourse structure, at least not for a full account of quantity implicatures.[1] On the other hand, we will single out a class of cases where the two theories make distinct predictions. In Sections 3 and 4 we present two inference task experiments that show (for a subclass of those cases) that Gricean reasoning at higher levels of discourse structure makes more accurate predictions. That, in turn, requires higher-level QUDs as input, which provides a piece of evidence that QUD-based discourse structure is needed after all.

It turns out that the discourses on which the predictions of the two theories diverge involve sequences of sentences with contrastive topics. The present study will therefore also shed new light on that relatively well-studied phenomenon. A typical example of contrastive topic use is given in (2). The first sentence provides information about Klaus, the second sentence says something about Martha. In that sense, Klaus and Martha are topics of the respective sentences. In addition, Klaus and Martha stand in a kind of contrast relation to each other. In this particular example one could characterize this relation as being ascribed two contrasting properties: "ate 2 apples" vs. "ate 4 bananas". Thus, we say that Klaus and Martha are *contrastive topics* (subscript *CT*), whereas *two apples* and *four bananas*, which represent two alternative instantiations of the object of eating that hold for the topics at hand, are the corresponding *foci* (subscript *F*).

(2)   [Klaus]$_{CT}$ ate [2 apples]$_F$. [Martha]$_{CT}$ ate [4 bananas]$_F$.

In English and German a CT+F structure can be realized by an intonational contour with two connected peaks—a *(fall-)rise* for the contrastive topic and a *fall* for focus—the "hat contour" due to its resemblance to a hat. In many languages, including German, a CT+F structure can also be signalled by word order, which usually involves movement of the contrastive topic to the left periphery of the sentence (see Krifka, 2008, for more details). In this paper, we will look mostly at German contrastive topics and will exploit the possibilities of CT-marking by word order available in that language.

Contrastive topics were among the first phenomena to which the idea of QUD-based discourse structure was applied. Büring (2003) (drawing on Roberts' 1996 theory) proposed that the use of a CT-contour signals an *inquiry strategy* consisting of discourse questions and subquestions. For example in (2), the contrastive topic marking on *Klaus* signals that the sentence answers an (implicit) question about Klaus, e.g. *What did Klaus eat?*, which in turn is part of

---

[1]Of course, there are other kinds of evidence for (not necessarily QUD-based) discourse structure.

a strategy to address a more general question *Who ate what?* The strategy then consists of a series of subquestions of the form *What did X eat?*, e.g. *What did Klaus eat?*, *What did Martha eat?*, etc. The strategy can be represented as a discourse tree, cf. Figure 1.

Who ate what?

What did Klaus eat?   What did Martha eat?

[Klaus]$_{CT}$ ate [2 apples]$_F$.   [Martha]$_{CT}$ ate [4 bananas]$_F$.
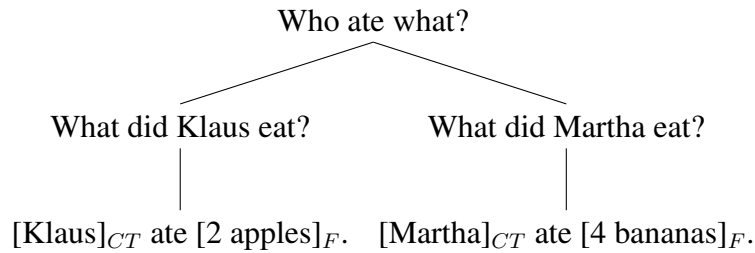
Figure 1: A discourse tree for (2).

Contrastive topics have been held responsible for the apparent disappearance of exhaustivity effects in answers to questions (Krifka, 1999; Hara and van Rooij, 2007; Büring, to appear). Under "normal" circumstances, i.e. if the focus accent in (3-B-a) is on *Mary* and there is no accent on *passed*, the answer is interpreted exhaustively like in (1): Mary and no one else passed the exam. However, if *Mary* bears a contrastive topic accent and the focus is on *passed*, cf. (3-B-b), then the answer is not exhaustive, but suggests on the contrary that there are other individuals besides Mary who might have passed or not. Büring (2003) and others have argued that a sentence with a contrastive topic, by convention, must give a strictly partial answer (i.e. non-exhaustive answer) to the main QUD (i.e. *Who ate what?* in figure 1 in the case of (2)).

(3)   A: Who passed the exam?
       B: a.   MARY$_F$ passed.
             b.   MARY$_{CT}$ PASSED$_F$.

In this paper (Section 5) we argue that in order to explain the full range of effects of contrastive topic on exhaustivity, the theory of contrastive topic must be backed up with a theory of discourse-level implicatures. In particular, our experiments show that the effect illustrated in (3) arises primarily in isolated sentences with contrastive topics. In sequences like (2) exhaustivity inferences are present, and they are due to discourse-level exhaustivity.

2   Two simple theories of discourse-level implicatures

After a brief introduction to sentence-level exhaustivity implicatures in Section 2.1, we present two possible approaches to discourse-level exhaustivity. Section 2.2 presents the *null theory*, where discourse-level pragmatic effects are derived from the plain conjunction of the implicatures of the individual sentences, i.e. independently of discourse structure. Section 2.3 presents a theory of the second type, where discourse-level implicatures are generated directly by applying Gricean reasoning to multi-sentence speech acts. Finally, Section 2.4 prepares the ground for an experimental investigation to compare the two theories.

2.1   Exhaustivity at sentence level

From a descriptive point of view, exhaustivity inferences are a class of inferences to the effect that *only* the individuals (or other semantic entities) mentioned explicitly in the sentence have

the property predicated of them, while the individuals not mentioned do not have that property, as in example (1) in Section 1: Venice is not mentioned and it was not visited by Cleo.

There is a variety of theoretical approaches to this phenomenon, ranging from predicate circumscription and minimal models (McCarthy, 1980; Groenendijk and Stokhof, 1984; Schulz and van Rooij, 2006) to approaches based on reasoning with pragmatic principles, such as Relevance Theory (Sperber and Wilson, 1986) and classical Gricean pragmatics. In this paper we use the latter in the way it has been employed in more recent literature (Sauerland, 2004; Geurts, 2010). In the Gricean picture, exhaustivity is an instance of *quantity implicature*—a class of implicatures in whose derivation the central role is played by the Maxim of Quantity:

(4)   Maxim of Quantity (1st submaxim):
        Make your contribution as informative as required.

Another important kind of quantity implicature is scalar implicature (*some* $\leadsto$ *not all*, *three* $\leadsto$ *not more than three*, etc.) The following example shows how the Maxim of Quantity is standardly employed to derive both kinds of implicatures.

(5)   A: What did Klaus eat?
        B: Klaus ate [2 apples and 4 bananas]$_F$.

Considering speaker A's question, speaker B could have uttered a number of more informative sentences, such as those in (6). If she believed, for instance, that in addition Klaus ate one pear (6-a), she would have said so, observing the Maxim of Quantity. However, she didn't. This could mean either that speaker B does not have any (positive or negative) belief about Klaus eating a pear, or that she believes that Klaus did not eat any pears. The first is excluded by the assumption that a speaker is competent, or opinionated on the question (this is usually assumed to be a precondition for a felicitous question-answer discourse). Hence we are left with the second possibility, that B believes that Klaus did not eat pears.

(6)   a. Klaus ate 2 apples, 4 bananas, and 1 pear.
        b. Klaus ate 3 apples and 4 bananas.
        c. Klaus ate 2 apples and 5 bananas.

In sum, this reasoning leads to the *exclusion of alternatives*. The exclusion of alternatives like (6-a) yields an exhaustivity implicature 'no other kinds of food', whereas the exclusion of (6-b) and (6-c) leads to the scalar implicatures 'not more than 2 apples', 'not more than 4 bananas'.

Finally, a couple of remarks on the nature of this set of alternatives. It is a closure under conjunction of the set of 'basic alternatives', which in turn is constrained in at least two ways. First, it is constrained by the *focus* of the actual utterance (5-B). The set includes only those alternatives that can be obtained from (5-B) by replacing the focused constituent *2 apples and 4 bananas* with alternative expressions of the same semantic type (Rooth, 1992): { 'Klaus ate 3 apples', 'Klaus ate some apples', 'Klaus ate small bananas', 'Klaus ate a pear', 'Klaus ate a steak', ... }. Second, it is constrained by the context and considerations of relevance. For instance, if speaker A is only interested in the question which fruit Klaus ate, and this is clear to B (no matter if the question has been explicitly asked or is understood from broader context), then the set of basic alternatives would not include, for instance, the alternative *Klaus ate a steak*, and its closure under conjunction, accordingly, would not include *Klaus ate 2 apples, 4 bananas, and a steak*. This means that information of steaks is not required and B does not violate Quantity if she does not mention them. Hence no implicature about steaks can be

drawn.[2]

A widely spread assumption is that the contextual restriction comes from the QUD. Questions can be seen as sets of non-exhaustive mutually compatible alternative answers along the lines of Hamblin (1973). The alternatives are generated from a *wh*-interrogative by replacing the *wh*-phrase by its possible instantiations, just like focus alternatives are generated by replacing the focused constituent. For example, the question *What did Klaus eat?* is represented as a set like {'Klaus ate 3 apples', 'Klaus ate some apples', 'Klaus ate small bananas', 'Klaus ate a pear', 'Klaus ate a steak', ...}. However, since the QUD by definition only contains relevant alternatives, i.e. for each proposition in the set it matters to the communication participants whether it is true or false, the set will not contain e.g. the alternative 'Klaus ate a steak' if we are only interested in fruit. Thus the set of alternatives used in Gricean reasoning contains only those focus alternatives that are also contained in the QUD. In the following section we will see how this basic recipe can be used on discourses that consist of more than one sentence.

2.2 Sentence-level implicature, and nothing else

Let us start with a *null theory* that consists just of the mechanisms described in the previous section—sentence-level Quantity implicatures, focus alternatives, and domain restriction of the focus alternatives by the QUD-alternatives at sentence level—and nothing else, and try to account for implicatures of units larger than a sentence. Just like (5), (7) answers the question *What did Klaus eat?*, but it expresses the same content in two sentences instead of one.

(7)   Klaus ate [2 apples]$_F$. He also ate [4 bananas]$_F$.

We first apply the reasoning schema presented in the previous section to calculate the implicatures of the first sentence, and get that Klaus did not eat anything but the two apples, i.e. he did not eat any pears or bananas in any quantity. The second sentence contradicts the part of the implicature concerning the bananas. However, Gricean conversational implicatures are cancelled if they get in conflict with the literal content of utterances. If implicatures are allowed to be cancelled partially, the second sentence in (7) cancels the part that Klaus did not eat 1, 2, 3 or 4 bananas, but leaves other parts intact, i.e. the implicature that Klaus did not eat any pears remains and the whole discourse in (7) carries this implicature, which is just an implicature of the first sentence that was not cancelled by the second.

Example (8) illustrates another way in which discourse-level pragmatic effects could result from sentence-level implicatures.

(8)   A: What did Klaus and Martha eat?
        B: Klaus ate [2 apples]$_F$. Martha ate [4 bananas]$_F$.

The first sentence of (8-B) has the implicature that Klaus did not eat anything except the two apples. Similarly, the second sentence implicates that Martha ate nothing but the four bananas. From the conjunction of these two implicatures one can derive that neither Klaus nor Martha ate, for instance, pears. Unlike in the previous example, this implicature does not result from either of the single sentences, but from the sum of the implicatures of both of them.[3]

---

[2]There are more constraints on alternative sets that have been proposed in the literature, including contrastivity (Zeevat, 2004; Büring, 2008), availability (Geurts, 2010) and relative complexity (Fox and Katzir, 2011), their role, however, will not be discussed in this paper.

[3]*Klaus* and *Martha* in (8) could also be marked as contrastive topics. However, this does not change anything in the analysis of this example assuming that only focus alternatives, and not topic alternatives are used in quantity implicature derivation.

The apparent discourse-level implicature in example (1) from Geurts (2010) can be given a similar account. Suppose the focus structure of Cleo's answer is as indicated in (9). Then the first sentence implicates, in particular, that Julius and Cleo did not go to Venice together. The second sentence implicates that Cleo did not go to Venice while Julius was in Milan. But if Cleo went to Venice neither with nor without Julius, then she did not go there at all. So the exhaustivity inference, that Cleo did not go to other places (for instance, Venice) follows from the conjunction of the exhaustivity implicatures of the individual sentences.

(9)   Julius and I first went to [Naples and Rome]$_F$ together.
        Then, while he went to see Milan, I visited [Ravenna]$_F$.

In other words, the null theory is sufficient to account for the discourse-level pragmatic effect in this example. The same applies to many other examples given by Geurts (2007, 2010) in support of Gricean reasoning at discourse level, which makes his argument largely inconclusive.

The null theory realizes the idea that discourse structure plays no role in the derivation of quantity implicatures of sequences of two or more sentences. However, is that generally the case? Consider the following example:

(10)   A: What did Klaus, Martha and Emma eat?
        B: Klaus ate [2 APPLES]$_F$. Martha ate [4 BANANAS]$_F$.

The only difference to (8) is that speaker A asks about three individuals, but one of them, Emma, is not explicitly mentioned in the answer. (Cf. also (3) in Section 1.) Does this discourse have an implicature that Emma did not eat anything? The null theory predicts no such implicature: If the focus is on the object DP, then the focus alternatives of the first sentence all have the form of *Klaus ate X*, where X runs through a set of edible things. Therefore the first sentence can only have an implicature that Klaus did not eat other things, but it cannot have an implicature about Emma. The same holds for the second sentence. Since neither set of focus alternatives includes alternatives about Emma, no implicature about Emma can be derived.

Speakers' judgments vary for this example: Some draw the implicature that Emma did not eat anything, some don't. We take it as an occasion for an experimental investigation. As will be shown in the next section, Gricean reasoning at discourse level does predict an implicature about Emma. In Section 2.4 we further show that examples like (10) involve a number of complications that might be the reason for the observed variability in the judgments. Concentrating on a more clear-cut case, we then compare the theories in an experimental study.

2.3 Discourse-level implicature and QUD structure

In the null theory presented in the previous section, an alternative is used in the derivation of an implicature only if it is both in the QUD, i.e. in the set of propositions that constitute the question in Hamblin's treatment, and in the set of focus alternatives of the sentence. The first change that we make is we allow the QUD-alternatives to be used in Gricean reasoning *directly* (11), whereas the relation to focus alternatives is indirect, via the QUD and the notion of question-answer congruence (12), cf. von Stechow (1991, p. 68).

(11)   The set of alternatives used in Gricean reasoning is the closure of the QUD under
        conjunction.

(12)   Question-Answer Congruence
        An answer is congruent to a QUD iff the set of QUD alternatives is the set of focus

6

alternatives of the answer, modulo constraints on the QUD.[4]

Question-answer congruence captures the intuition that focused (accented) constituents in the answer must correspond to *wh*-phrases in the question. It is easy to see that so far the theory makes the same predictions for simple cases like (5) as the null theory. If Gricean reasoning is applied to QUD alternatives, but those are the same as the focus alternatives, this leads to the same result as if one applied Gricean reasoning to focus alternatives directly.

Second, we assume that discourse is hierarchically structured by QUDs, as illustrated in Figure 1 in Section 1. If a QUD is answered by more than one sentence, it is split into a series of subquestions, which constitute the *strategy of inquiry* to address it, such that each sentence of the answer addresses one of the subquestions. For example, in (8), repeated below, the speaker adopts a strategy to address the main question *What did Klaus and Martha eat?* by splitting it into two subquestions, one about Klaus and one about Martha. In strategies of the kind that we consider in this paper, the alternative set of the subquestion is always a proper subset of the alternative set of the main question. For instance, the set of QUD alternatives of the main question in (13) contains alternatives of the form { Klaus ate 1, 2, 3, ... apples, Klaus ate 1, 2, 3, ... bananas, Klaus ate ..., Martha ate 1, 2, 3, ... apples, Martha ate 1, 2, 3, ... bananas, Marta ate ... },[5] while the QUD that immediately dominates the first sentence is the subset containing only the alternatives about Klaus.

(13)   What did Klaus and Martha eat?
           What did Klaus eat? Klaus ate [2 apples]$_F$.
           What did Martha eat? Martha ate [4 bananas]$_F$.

The QUD that immediately dominates a particular single utterance (or sentence) is called its *local* QUD (or the congruent QUD, see Simons et al., this issue). Notice that the first sentence *Klaus ate [2 apples]$_F$* is only congruent with its local QUD *What did Klaus eat?*, but not with the overarching QUD *What did Klaus and Martha eat?* This is the reason why this sentence with the given information structure can only be a proper part of a strategy to answer that QUD, but it cannot be the whole answer. Crucially, the notion of question-answer congruence only applies to local QUDs, because focus-background structure is only defined for single utterances (or sentences, or clauses), and it is less clear what could be *the* focused constituent, and hence *the* set of focus alternatives, of two or more sentences.

---

[4]"Modulo constraints on the QUD" means that the QUD can actually be a proper subset of the set of focus alternatives, containing only the currently relevant alternatives (domain restriction), only available alternatives, and only alternatives that are not syntactically more complex than the answer (cf. fn. 2), whereas the focus alternatives are generated in a context-independent way by replacing the focused constituent by *all* alternative expressions of the same semantic type (Rooth, 1992). However, to satisfy congruence the QUD must have the same variables as the open proposition derived from focus structure. For example, if a sentence has VP focus and its set of focus alternatives is of the form { Mary ate an apple, Mary ate a banana, Mary peeled an apple, Mary went for a walk, ... }, then the QUD may not consist just of the propositions { Mary ate an apple, Mary ate a banana }, because these alternatives vary just in the instantiation of the object of eating, and not of the whole VP.

It is more common in the literature to make the set of focus alternatives directly sensitive to all those constraints (see e.g. Fox and Katzir, 2011). In the present theory, it does not matter whether the focus alternatives are constrained in this way, but it does matter that the QUD is constrained.

[5]The question *What did Klaus and Martha eat?* is ambiguous between a reading where the speaker wants to know what the group of individuals, consisting of Klaus and Martha, ate, and a reading where for each of Klaus and Martha, the speaker wants to know what he or she ate (cf. Krifka, 2001). In the first case, the set of QUD alternatives follows the pattern: *Klaus and Martha ate 2 apples, Klaus and Martha ate 4 bananas, Klaus and Martha ate ...,* etc. In the second case alternatives of the from *Klaus ate X* and *Martha ate X* are also included. The latter is equivalent to the set of alternatives of the question *Who ate what?* where the *who*-variable is restricted to the set of Klaus and Martha.

Finally, the third point in which we depart from the null theory:

(14)   Not only the local QUDs, but also higher level QUDs participate in the derivation of Gricean implicatures.

In other words, Gricean quantity implicatures are first calculated for each individual sentence with respect to the set of alternatives of its local QUD. Then the implicatures of the sequence of sentences are calculated with respect to the QUD that dominates the whole sequence.

Applying this to (15), we first calculate the implicatures of each individual sentence and we get implicatures about Klaus (Klaus did not eat anything but two apples) and Martha (Martha did not eat anything but four bananas), but we get no implicature about Emma. However, the main QUD *What did Klaus, Martha and Emma eat?* does include alternatives about Emma in addition to those about Klaus and Martha. So during the interpretation of (15), the hearer should be wondering why the speaker did not make a stronger statement, for instance, that Klaus ate two apples, Martha ate four bananas and Emma ate a pear, which would have been relevant given the QUD. He has to conclude that the speaker does not believe that this is the case, and given that a speaker does believe what she claims explicitly, that is, that Klaus ate two apples and Martha ate four bananas, this means that the part the speaker does not believe is that Emma ate one pear. Finally, on the assumption that the speaker is opinionated, the hearer concludes that the speaker believes that Emma did not eat a pear. Applying the same reasoning to other alternatives of the form *Emma ate X* we derive the implicature that Emma did not eat anything.

(15)   What did Klaus, Martha, and Emma eat?
          What did Klaus eat? Klaus ate [2 apples]$_F$.
          What did Martha eat? Martha ate [4 bananas]$_F$.

Notice that this is a genuine discourse-level implicature of the kind envisaged by Geurts (2007, 2010) that goes beyond the mere sum of sentence-level implicatures. However, is this a correct prediction? Does the answer in (10) have such an implicature? As we already mentioned in Section 2.2, judgments vary on this point. The preliminaries for an experimental investigation of this question are laid out in the next section.

2.4   Towards an experimental investigation

The empirical question that we want to investigate is whether there are discourse level implicatures beyond the sum of sentence level implicatures, that is, whether we need a mechanism to generate implicatures above the sentence level such as the one presented in the previous section, or whether no such mechanism is needed and the null theory from Section 2.2 is sufficient.

In previous sections we have seen that the two theories make different predictions in examples like (10), where the question asks about three individuals, but the answer only mentions two of them. However, this particular example and similar cases where the topic of the question and the focus in the answer correspond to main constituents of the sentence (subject and direct object, respectively, in (10)) present a complication which makes them less suitable as a model for critical stimuli in an experimental study: It is generally accepted that a question like *What did Klaus eat?* can normally only be asked in a situation where it is common ground that Klaus ate something. Similarly, the question *What did Klaus, Martha, and Emma eat?* seems to "presuppose" that each of the three individuals ate something. However, this preempts the potential implicature that Emma ate nothing. Of course, one can felicitously reply ... *and Emma ate nothing*, but if the statement about Emma is not made explicitly, the corresponding implicature does not seem strong enough to defeat the assumption associated with the question. This could

be the reason why judgments vary for this example. Furthermore, according to our intuitions the interrogative *Who (of Klaus, Martha, and Emma) ate what?*, which would be another way to make the relevant QUD explicit, gives rise to the same problem. Even though by the same reasoning the double *wh*-question should only require a shared assumption that someone ate something, in practice the question is typically asked only if there is a known set of people who ate something, and a set of food items eaten, and the question only asks to match the eaters with the food. This seems sufficient to trigger the expectation that Emma ate something too. *What happened?* is the only question we could think of that does not preempt the implicature about Emma and which could be felicitously answered by *Klaus ate 2 apples. Martha ate 4 bananas.* However, in that case one would normally assume broad focus on the whole sentences, which gives rise to a broader set of focus alternatives and makes it possible to derive the implicature about Emma in the null theory following the same reasoning as in example (7).

One way to avoid this problem is to consider cases where the topic and the focus are sub-constituents of a single phrase, e.g. a DP. A construction that fulfills this requirement is found, for instance, in German sentences with *split topicalization*, illustrated in (16).

(16)   ÄPFEL$_1$ hat   er   [ZWEI t_1]$_{DP}$ gegessen.
       apples   has   he   two   eaten

Split topicalization is characterized by movement of a part of a phrase to the left periphery of the sentence. In (16) the underlying determiner phrase (DP) is *zwei Äpfel* 'two apples', where the noun *Äpfel* is fronted, while the determiner *zwei* 'two' stays stranded in its underlying position (De Kuthy, 2002; Fanselow and Ćavar, 2002). Sentences with split topicalization exclude the possibility of a broad focus reading, that is, they normally cannot be uttered out of the blue or serve as answers to questions like *What happened?* Moreover, De Kuthy (2002) argues that the two parts of the split phrase may not be part of the same information-structural projection. This means basically that in (16) the numeral must bear narrow focus (accent on the determiner does not project to the level of the DP, cf. Selkirk, 1995), and the fronted noun is a contrastive topic.[6]

From a theoretical point of view, (17) is analogous in all relevant respects to (10). Since both sentences have narrow focus on the numeral, they are congruent answers to the questions *How many apples did Klaus eat?* and *How many bananas did Klaus eat?*. Their focus alternatives include alternative quantities of apples and bananas, but no alternatives about pears. Therefore, the null theory predicts the scalar implicatures 'Klaus did not eat three, four, five, etc. apples' and 'Klaus did not eat five, six, etc. bananas' for the first and the second sentence, respectively, but it does not predict an implicature that Klaus did not eat pears. In contrast, the overarching QUD *What did Klaus eat?* can be assumed to include alternatives ranging over different quantities of all three kinds of fruit: Klaus ate X apples, Klaus ate Y bananas, Klaus ate Z pears. Therefore, the QUD-based theory predicts the inference that Klaus did not eat pears.

(17)   A: Was (von den Äpfeln, Bananen und Birnen) hat Klaus gegessen?
            'What (of the apples, bananas, and pears) did Klaus eat?'
       B: [ÄPFEL   ]$_{CT}$ hat   er   [ZWEI   ]$_F$ gegessen.
                apples   has   he   two   eaten

---

[6]Another theoretical possibility is that the fronted noun bears a second narrow focus, giving rise to a double focus construction. In that case, both the fruit type variable and the number of pieces variable participate in the generation of focus alternatives at sentence level, so a sentence like (16) is predicted to implicate that Klaus did not eat any other number of any other fruit (in particular, no pears) even in the null theory. Intonation can disambiguate between CT+F and F+F structures, but it was absent in our experiments because we used written stimuli. Nevertheless, our final results suggest that CT+F must have been the preferred reading in our experimental materials, cf. discussion in Fn. 8, Section 3.4.

[BANANEN]$_{CT}$ hat er [VIER]$_F$ gegessen.
bananas   has he   four   eaten
'As for apples, he ate two. As for bananas, he ate four.'

Examples like (17) have a number of advantages over (10). First, the question *What did Klaus eat?* only requires that Klaus ate something, but not that he ate a certain positive amount of every relevant kind of fruit. The trick here is that the topic and the focus phrase in the answer in (17) do not correspond to two distinct *wh*-variables in the question, but to a single *wh*-phrase, which they "split" between themselves. Second, we avoid the "broad focus" problem that arises with *What happened?* questions in (10) by using the marked word order as explained above. Finally, the word order signals that apples and bananas are contrastive topics, which (in Büring's theory of CT) is an explicit indicator that a QUD strategy is involved and therefore the explicit question *What did Klaus eat?* is not terminal but must be in a higher structural position. If we can show that speakers compute exhaustivity implicatures using the alternative set of that question, we show that higher-level QUDs are involved in Gricean reasoning.

As one control condition we take a case where both theories predict an inference like 'Klaus did not eat any pears', e.g. a one-sentence answer without topicalization (cf. (5) in Section 2.1):

(18)   A: Was (von den Äpfeln, Bananen und Birnen) hat Klaus gegessen?
             'What (of the apples, bananas, and pears) did Klaus eat?'
        B: Er hat [zwei ÄPFEL und vier BANANEN]$_F$ gegessen.
             he has two apples and four bananas   eaten
             'He ate two apples and four bananas.'

If the null theory is right, we should expect to see less 'Klaus did not eat any pears' inferences in (17) than in (18). If the QUD-based theory holds in its strongest version, that is, if implicatures at higher levels of discourse structure are drawn as frequently and reliably as sentence-level implicatures, then we should expect the number of 'Klaus did not eat any pears' inferences in (17) and (18) to be equally high.

As a second control condition we look at cases where we do not expect an exhaustivity inference under either theory, namely single sentences with contrastive topics. (cf. ex (3), sec. 1):

(19)   A: Was (von den Äpfeln, Bananen und Birnen) hat Klaus gegessen?
             What (of the apples, bananas, and pears) did Klaus eat?
        B: [ÄPFEL]$_{CT}$ hat er [ZWEI]$_F$ gegessen.
             apples   has he two   eaten
             'As for apples, he ate two.'

The null theory predicts no exhaustivity inference about pears or other food for the same reasons as in (17). Therefore on the null theory, the number of 'Klaus did not eat any pears'-inferences in (17) and (19) should be equally low.

The predictions of the QUD-based theory depend here on our assumptions about the function of contrastive topic (CT). Non-exhaustivity can either be associated with CT by convention along the lines of Krifka (1999) and Hara and van Rooij (2007), cf. Section 1, or one can adopt a weaker theory such as Büring's (2003), where CT just indicates a QUD strategy of the form *How many apples did Klaus eat? How many bananas did Klaus eat? How many X did Klaus eat?*, etc., while non-exhaustivity would follow from Gricean reasoning. The reasoning in (19) would go roughly as follows: Speaker B could have uttered (18-B) either believing that Klaus ate two apples and nothing else or believing that he ate two apples and something else. But if the speaker believed that Klaus ate nothing but the two apples (the exhaustive reading), would

she have uttered (18-B)? Probably not, because a sentence *Er hat zwei Äpfel gegessen* 'He ate two apples' without a CT would have been a less effortful answer (see Gricean Maxim of Manner) being otherwise just as informative. Normally, it is not rational to pursue a complex inquiry strategy as the one indicated by the CT in (19) if one has positive information about only one topic. Therefore, speaker B, assuming that she is rational, could not have meant that Klaus ate two apples and nothing else. Or put more simply, CT indicates that the speaker is planning to discuss another topic, hence exhaustivization with respect to the overarching QUD is suspended until the discussion of all its subtopics can be safely assumed to be complete. In contrast, the use of a CT in (17) is rationally justified because there are two distinct topics— apples and bananas—about which the speaker has something to say. The speaker could have uttered (17) in the belief state where Klaus ate nothing but the two apples and the four bananas. To exclude the pears the standard Quantity-based reasoning is applied.

In other words, with a suitable theory of CT in the background, the QUD-based approach makes distinct predictions for (17) and (19). We expect less exhaustivity inferences as 'Klaus did not eat any pears' in the latter case. Predictions of both theories are summarized in Table 1.

	'Klaus did not eat any pears'	
	null theory	QUD-based
Er hat 2 Äpfel und 4 Bananen gegessen.	+	+
Äpfel hat er 2 gegessen. Bananen hat er 4 gegessen.	−	+
Äpfel hat er 2 gegessen.	−	−

Table 1: Predictions of the null theory and the QUD-based theory with respect to presence (+) vs. absence (−) of the inference 'Klaus did not eat any pears'.

3 Experiment 1

3.1 Predictions

This experiment addresses the question whether higher-level QUDs participate in the generation of exhaustivity implicatures. The previous section presented the predictions of the null and the QUD-based theory for a range of examples which can be used as a model for critical and control stimuli to help resolve this question. To avoid a potential issue concerning the obligatory presence of a conjunction in single sentences vs. a preference against conjunction in sentence pairs (cf. table 2), we first look at pairs of conjoined clauses with a characteristic ellipsis pattern, cf. (20), instead of two full sentences separated by a full stop.

(20)   Äpfel hat er zwei ~~gegessen~~ und Bananen ~~hat er~~ vier gegessen.
        apples has he two eaten   and bananas has he four eaten.

Since the CT+F structure, but not the presence of a full stop, is relevant for the QUD-based theory, its predictions for (20) are the same as for (17) in Section 2.4.

3.2 Method

Subjects: Thirty-three self-reported native speakers of German (6 male, age 19-55 years, mean 24 years, all students) took part. We raffled 3 times €25 among the participants.

Materials: We adopted the inference task experimental paradigm previously employed in the study of scalar implicatures (Geurts and Pouscoulous, 2009; Chemla and Spector, 2011). A typical experimental item is given in (21).

(21)   A: In der Schale lagen Äpfel, Bananen und Birnen. Klaus war sehr hungrig.
             in the bowl   lay   apples bananas and pears   Klaus was very hungry

       B: Was hat er gegessen?
             what has he eaten

       A: Äpfel hat er zwei und Bananen vier gegessen.
             apples has he two   and bananas four eaten

Items contained a short dialogue between two speakers (A and B in (21)). The first sentence of speaker A introduced the relevant alternatives to be taken into account in pragmatic interpretation of the critical sentence, e.g. *There were apples, bananas, and pears in a bowl*.[7] The second sentence introduced the "protagonist" of the story (Klaus) and created the context for the subsequent question by speaker B. The question targeted the object phrase of a transitive verb, e.g. *What did he eat?*, followed by the answer of speaker A—the critical sentence of the item. The critical sentences varied in four conditions in a 2×2 design: the number of ALTERNATIVES mentioned in the sentence (1 vs. 2) and the presence vs. absence of TOPICALIZATION:

(22)   a.   Er hat zwei Äpfel   gegessen.   1ALT,−TOP
             he has two   apples   eaten
       b.   Äpfel hat er zwei gegessen.   1ALT,+TOP
             apples has he two   eaten
       c.   Er hat zwei Äpfel   und vier Bananen gegessen.   2ALT,−TOP
             he has two   apples   and four bananas   eaten
       d.   Äpfel hat er zwei und Bananen vier gegessen.   2ALT,+TOP
             apples has he two   and bananas   four eaten

In the 1ALT condition, the target sentence mentioned only one of the alternatives listed in the first sentence of the context, cf. (21), e.g. apples. In the 2ALT condition, two such alternatives were mentioned, e.g. apples and bananas. They were presented as coordinate DPs (in 2ALT,−TOP) or as elliptical coordinate clauses (in 2ALT,+TOP) connected by the conjunction *und* 'and'. The first sentence of the context also varied along the ALTERNATIVES dimension:

(23)   a.   In der Schale lagen Äpfel   und Birnen.   1ALT
             in the bowl   lay   apples and pears
       b.   In der Schale lagen Äpfel, Bananen und Birnen.   2ALT
             in the bowl   lay   apples bananas   and pears

It always listed one extra alternative, e.g. *Birnen* 'pears', i.e. (22-a) and (22-b) would go with (23-a), whereas (22-c) and (22-d) with (23-b). This extra alternative was the one targeted by the inference task. The task followed the dialogue and was formulated as in (24).

---

[7]Fisseni (2010) tried to elicit exhaustivity inferences in a number of experimental designs. In general, the study suggests that exhaustivity is difficult to elicit. However, the designs where the alternatives for exhaustivity were explicitly given measured relatively higher numbers of exhaustivity inferences. With this potential problem in mind, we tried to maximize our chances to detect any exhaustivity inferences at all. This is why we introduced the alternative sets explicitly in the context, and this is also one of the reasons for using the inference task paradigm, which has been argued to push up the numbers of quantity implicatures (see Geurts and Pouscoulous, 2009).

(24)　　a.　Würden Sie daraus　folgendes schließen:
　　　　　　would　you from.this following infer
　　　　　　'Would you infer the following from this:'
　　　　b.　Klaus hat keine Birnen gegessen.
　　　　　　Klaus has no　pears　eaten
　　　　　　'Klaus did not eat any pears.'

We also balanced the type of phrases that underwent split topicalization. The items were equally distributed among the three groups: numeral+noun DPs as in (22-d); pseudopartitives consisting of a numeral, a measure unit and a mass noun, with mass nouns undergoing topicalization, numeral+measure unit stranded, as in (25); and participial VPs containing a quantificational adverbial like *dreimal* 'three times', participle head topicalized, adverbial stranded, as in (26).

(25)　　Bier hat er zwei Flaschen und Rotwein drei　Gläser getrunken.
　　　　Beer has he two　bottles　　and red.wine three glasses drunk
　　　　'He drank two bottles of beer and three glasses of red wine.'

(26)　　Gegrillt　hat er dreimal　　und gebadet zweimal.
　　　　barbecued has he three.times and bathed　two.times
　　　　'He had a barbecue three times and went swimming two times.'

Procedure:　　The experiment was conducted as web-experiment using *WebExp* (Keller et al., 2009). After a training phase of three training items, each participant saw six items per condition, adding up to 24 experimental items and 48 additional fillers, i.e. 72 stimuli per participant. We created four lists of items in a Latin square design. In each list each item occurred in only one condition. The lists were automatically randomized by the WebExp software.

　　The participants had to choose between two responses *Ja, würde ich* 'Yes, I would' and *Nein, würde ich nicht* 'No, I would not' by mouse-clicking on the response buttons at the bottom of the screen with the positive answer on the left and the negative answer on the right. The dialogue (21), the task (24), and the buttons were presented simultaneously on the screen. There was no time limit in answering the questions. The experiment took approximately 20 minutes.

3.3　Results

Prior to the analyses we checked if the participants were paying attention to the task by looking at their responses to filler items, which were meant to allow for only one possible inference. We first excluded the fillers that turned out to be more ambiguous than we had expected, that is, if the fillers received more than five unexpected responses. All subjects (two in total) that gave more than three unexpected answers to the remaining fillers were excluded from the analysis. Due to a coding error we also had to exclude one experimental item. Consequently, data from 31 subjects and 23 experimental items entered the statistical analysis.

　　We fitted logistic mixed-effects models using the *glmer*-function of the *lme4*-package (Bates et al., 2013) of the statistics software *R* (R Core Team, 2013). With model comparisons we identified the best model using the Answer as predictor variable, including the experimental factors ALTERNATIVES and TOPICALIZATION as fixed effects. We also included the centered value of Trial (position of the stimulus in the experiment) as fixed effects to control for effects of training or fatigue that might change the reasoning behaviour. In the random effect structure we included random intercepts per subjects and random slopes of TOPICALIZATION per subjects.

　　We found significant main effects of TOPICALIZATION ($\beta = -5.22$, $z = -6.00$, $p < 0.001$) and centered Trial ($\beta = 0.014$, $z = 1.99$, $p = 0.047$), meaning that the probability of giving a

yes-response was lower for topicalized sentences and the general probability of answering *yes* increased over the course of the experiment. We also found an interaction of ALTERNATIVES and TOPICALIZATION ($\beta = 4.02$, $z = 5.55$, $p < 0.001$), meaning that we received significantly more *yes*-responses for topicalized sentences in the 2ALT-condition than in the 1ALT-condition. The absolute number and percentages of the *yes*-responses are displayed in table 2.

Condition	Number of *yes*-Answers	Total Answers	Percentage of *yes*-Answers
1ALT,−TOP	157	178	88.20%
1ALT,+TOP	51	178	28.65%
2ALT,−TOP	163	179	91.06%
2ALT,+TOP	129	178	72.47%

Table 2: Summary of *yes*-answers per condition in experiment 1

3.4 Discussion

The results show a high number of *yes*-responses—88.20% and 91.06%—in the control conditions without topicalization, 1ALT,−TOP and 2ALT,−TOP, respectively. That is, in nine out of ten cases, participants agreed that sentences like the German counterparts of *He ate two apples* and *He ate two apples and four bananas* implied that 'he' did not eat pears. This supports the standard assumption that exhaustivity inferences are normally drawn in direct answers to *wh*-questions. The significant main effect of TOPICALIZATION shows that speakers are generally less likely to draw exhaustivity inferences in sentences with topicalization. This, again, is in line with the theories of contrastive topic such as Büring (2003), Krifka (1999), and others, which associate contrastive topic with partial answerhood and non-exhaustivity.

In the experimental condition 2ALT,+TOP, participants agreed that the target sentence had an exhaustivity implicature in 72.47% of the cases. In other words, in a vast majority of cases participants agreed that a sentence like *Äpfel hat er zwei und Bananen vier gegessen* 'As for apples, he ate two; as for Bananas, he ate four' implied that 'he' did not eat any pears. The most important result is that this inference is drawn significantly more often than in the 1ALT,+TOP condition, i.e. in sentences with topicalization that consist of a single clause and name only one alternative *Äpfel hat er zwei gegessen* 'As for apples, he ate two'—cf. 28.65% *yes*-responses. If the null theory were right we should have found equally low numbers of *yes*-responses in both +TOP conditions. The fact that 2ALT,+TOP is significantly "more exhaustive" than 1ALT,+TOP in our data speaks in favour of the QUD-based theory of discourse-level implicatures.[8]

At the same time the percentage of exhaustivity inferences in 2ALT,+TOP was significantly lower than in the control conditions without topicalization: 72.47% vs. 88.20% and 91.06%. This is against the predictions of the QUD-based theory in its strongest version, which says that discourse-level implicatures are drawn *always*, i.e. whenever and as frequently as sentence-level

---

[8]28.65% might still be seen as a rather high number given that the theories predict no exhaustivity at all in this condition. A possible explanation is that in a certain percentage of the cases the participants could have construed the fronted expression in the target sentences as a focus rather than a contrastive topic, see fn. 6 in Section 2.4 on the double focus reading. At the same time, the fact that we found any effects at all suggests that the participants did not uniformly assume a double focus structure, which would have led to an equally high number of *yes*-responses to the exhaustivity inference in all four conditions. In other words, the participants must have assumed the CT+F reading in the +TOP condition frequently enough, so the fact that we used written materials without intonation (which would have disambiguated between F+F and CT+F structures) was not a problem.

implicatures are drawn. However, this result is consistent with a milder version of the QUD-based theory which allows for additional factors that can impede the derivation of implicatures at discourse level in some cases. One such factor could be the extra effort processing a split topicalization construction and deriving discourse-level implicatures in addition to the sentence-level ones. The fact that these potential factors did not kill the effect altogether suggests that the participants were trying hard to draw discourse-level implicatures in this experiment.

However, the observed pattern allows for an alternative explanation. What if the null theory is right and accounts for the main effect of TOPICALIZATION and the difference between the 72.47% 'yes' in the 2ALT,+TOP vs. the 88.20% and 91.06% 'yes' in the −TOP conditions, whereas the difference between 2ALT,+TOP and 1ALT,+TOP (28.65% 'yes') is due to some independent factor, e.g. world knowledge? The problem with our manipulation of ALTERNATIVES was that the target sentences in the 1ALT and the 2ALT conditions systematically had distinct propositional content. Differences in propositional content may lead to differences in implicatures, independently of exhaustivity. Two apples is not much food for a hungry person, so one might infer that Klaus must certainly have eaten something else. In contrast, two apples and four bananas is a lot of food, so it is plausible to infer that Klaus did not eat anything else. This would be an inference based on world knowledge rather than exhaustivity and it could account for the higher number of 'no pears' inferences in the 2ALT,+TOP, and more generally for the main effect of ALTERNATIVES. Experiment 2 addresses this possibility.

4   Experiment 2

4.1   Predictions

This experiment was designed to replicate the results of Experiment 1 excluding potential confounds. The first concern was that the target sentences in 1ALT and 2ALT conditions represented different propositional content yielding different implicatures based on world knowledge (see discussion in Section 3.4). We replaced the 1ALT condition by unfinished lists like (27):

(27)   Äpfel hat er zwei gegessen, Bananen hat er vier gegessen, und . . .
       apples has he two   eaten   bananas has he four eaten   and

As in the 1ALT,+TOP condition in Experiment 1, no exhaustivity inference is expected here. In any Gricean theory, i.e. in our null theory and in the QUD-based theory in particular, exhaustivity implicatures result from reasoning about what the speaker did not say (as explained in Section 2.1). The inference that Klaus did not eat pears could only be made if the hearer has sufficient evidence, e.g. a full stop or falling "full stop intonation" on bananas, that the speaker is not going to say *two apples, four bananas and one pear*. In contrast, *"and ..."* indicates that the speaker might still add something relevant to the answer. Therefore the information on what the speaker did not say serving as input for Quantity-based pragmatic reasoning is not yet available, and the exhaustivity implicature cannot be drawn. With this change in the design, the propositional content was the same in all conditions.

The second concern had to do with the fact that we used conjoined clauses rather than proper sentences separated by a full stop in the 2ALT conditions of Experiment 1 (see remarks on (20) in Section 3.1). Although the predictions of the QUD-based theory should not depend on this difference, it is less clear what the null theory has to say in the case of conjoined clauses. One might be able to formulate a version of the null theory that would derive the 'no pears' implicature in sentences like (20) just from the focus structure of this complex sentence, without

reference to QUD-based discourse structure. In other words, we cannot be sure that the 'no pears' implicature in (20) is a discourse-level implicature in the strict sense, because the whole "discourse" consists of one sentence. In Experiment 2 we therefore replaced the conjoined clauses by two full sentences with contrastive topics separated by a full stop, i.e. the kind of sequences we originally started in Section 2.4:

(28)   Äpfel hat er zwei gegessen. Bananen hat er vier gegessen.
        apples has he two eaten   bananas has he four eaten

A question that arises in this connection is whether the two sentences should be connected by a conjunction *und* 'and', as in the control condition without topicalization (where the conjunction is obligatory), or whether the sentences should be juxtaposed, as in (28) and as in the 3DOTS condition in (27). A few things can go wrong with conjunction. Sentence-initial conjunction is considered bad style in written language. Moreover, in languages like English and German, if *and* conjoins three or more elements, the conjunction typically appears only before the last element of the list. Therefore it may serve as a signal that the list is about to be completed, which could affect exhaustivity inferences. Therefore, we constructed two versions of the critical condition, one with and one without the conjunction to evaluate its potential effect.

4.2 Method

Subjects: Thirty-five self-declared native speakers of German (8 male, age 18-37, mean 20, all students) took part. We raffled 3 times €25 among 29 participants; 6 further participants were paid €7 each.

Materials: The items in this experiment were identical to those in Experiment 1, cf. (21) and (24) in Section 3.2, except for the critical sentences and the first context sentence. The critical sentences varied in four conditions:

(29) a. Er hat zwei Äpfel und vier Bananen gegessen.   1SENT
        he has two apples and four bananas eaten

     b. Äpfel hat er zwei gegessen. Und Bananen hat er vier gegessen.   2SENT,+UND
        apples has he two eaten   and bananas has he four eaten

     c. Äpfel hat er zwei gegessen. Bananen hat er vier gegessen.   2SENT,−UND
        apples has he two eaten   bananas has he four eaten

     d. Äpfel hat er zwei gegessen, Bananen hat er drei gegessen und . . .   3DOTS
        apples has he two eaten   bananas has he three eaten   and

Condition 1SENT (29-a) consisted of one declarative sentence in canonical word order with an internal coordination *und* 'and' and was identical to the condition 2ALT,−TOP of Experiment 1, cf. (22-c). Both 2SENT conditions consisted of two sentences with split topicalization. In the 2SENT,+UND condition (29-b), the second sentence started with the conjunction *und* 'and'; in 2SENT,−UND (29-c) there was no conjunction. The condition 3DOTS (29-d) consisted of two clauses with split topicalization, separated by a comma, followed by the conjunction *und* 'and' and three dots. The second difference to Experiment 1 was that the first sentence of the context invariably introduced three alternatives, e.g. *There were apples, bananas, and pears in a bowl*, as in the 2ALT conditions of Experiment 1, cf. (23-b).

Procedure: The procedure was largely the same as in Experiment 1. The experiment was conducted as a web-experiment using the *php*-based online tool *OnExp* (http://onexp.textstrukturen. uni-goettingen.de/). After the training phase of three filler items, each participant saw six items per condition, i.e. 24 items in total and additional 64 fillers, summing up to 88 stimuli. Again, we provided four different lists of items in a Latin square design. The lists were automatically randomized for each participant by the OnExp software. The experiment took about 25 minutes.

The fillers were designed to show a great variety in terms of difficulty and type of inference (logical, pragmatic, world knowledge based, etc.) to avoid potential bottom or floor effects. To mask the experiment material sufficiently, we also included fillers ending in three dots. In many fillers none of the responses could have been characterized as "right" or "wrong". Therefore it did not make sense to exclude any participant on the basis of their performance on fillers, as we did in Experiment 1. However we had to exclude one participant because she also participated in Experiment 1. Therefore data from 34 subjects and 24 items entered the statistical analysis.

4.3 Results

For the analysis we fitted a logistic mixed-effects model using the *glmer*-function of the *lme4*-package (Bates et al., 2013) of the statistics software *R* (R Core Team, 2013).

With model comparisons we identified the best model using the Answer as predictor variable, including the experimental conditions as fixed effects. In the random effect structure we included random intercepts per subjects and random intercepts per items.

We found significant main effects of the following conditions: 1SENT received the most *yes*-responses of all four conditions and was therefore set as baseline for group comparisons. 2SENT,+UND ($\beta = -1.132$, $z = -2.36$, $p = .0184$ and 2SENT,−UND ($\beta = -1.429$, $z = -3.06$, $p = .0022$) received significantly less *yes*-responses than 1SENT. 3DOTS received the fewest *yes*-responses, significantly less than 1SENT ($\beta = -7.484$, $z = -11.30$, $p < .0001$).

Post-hoc test with Pearson's χ^2-test revealed that the difference between 2SENT,+UND and 2SENT,−UND is not significant ($\chi^2(1) = 0.4954$, $p = .48$). The difference between 2SENT,−UND and 3DOTS on the other hand is significant ($\chi^2(1) = 193.1$, $p < .0001$). The absolute number and percentage of *yes*-responses are displayed in table 3.

Condition	Absolute *yes*-Answers	Total Answers	Percentage of *yes*-Answers
1SENT	190	204	93.14%
2SENT+UND	177	204	86.76%
2SENT-UND	172	204	84.31%
3DOTS	32	204	15.69%

Table 3: Summary of *yes*-answers per condition in experiment 2

4.4 Discussion

The results of this experiment confirm the findings of experiment 1. First, the control condition 1SENT (*He ate two apples and four bananas*) led to roughly the same percentage of exhaustivity inferences (93.14%) as the same sentences in experiment 1, cf. 91.06% *yes*-responses in the 2ALT,−TOP condition. This finding is robust across the two experiments.

Second, the number of exhaustivity inferences in the two experimental conditions 2SENT,+UND and 2SENT,−UND was again high, 86.76% and 84.31%, respectively (cf. 72.47% in 2ALT,+TOP

in experiment 1). These were the conditions where, by our hypothesis, exhaustivity inferences could only arise from pragmatic reasoning at discourse level, with the alternatives of the higher-level QUD. The difference between the two experiments was in whether the target sequence consisted of two elliptical clauses conjoined by *und* 'and' (*Äpfel hat er zwei und Bananen vier gegessen*), or of two sentences separated by a full stop, with or without *und* (*Äpfel hat er zwei gegessen. (Und) Bananen hat er vier gegessen*). The results suggest that discourse-level exhaustivity inferences are drawn in both cases. If in the first experiment one could doubt whether the inferences were truly discourse-level because the target unit consisted of, strictly speaking, just one sentence, experiment 2 puts these doubts out of the way. The presence vs. absence of the conjunction *und*, in turn, could not be shown to play a role.

Finally, the number of exhaustivity inferences in both experimental conditions was significantly higher than in the second control condition 3DOTS (*Äpfel hat er zwei gegessen, Bananen hat er vier gegessen, und...*) where no exhaustivity inference was expected due to the explicit cue *und...* indicating that the discourse is not finished. The number of *yes*-responses to the inference 'Klaus did not eat any pears' in this condition was as low as 15.69%. In contrast to the first experiment, this difference cannot be attributed to an inference based on world knowledge since the target units in all conditions now had identical propositional content. That is, in all conditions Klaus ate the same amount of food as far as the literal content of the target units is concerned, so there is no reason why it would be more plausible in one condition than in another that Klaus ate or didn't eat something else. This leaves us with the QUD-based explanation of the observed pattern: the higher number of exhaustivity inferences in a sequence like *Äpfel hat er zwei gegessen. (Und) Bananen hat er vier gegessen.* must be due to a discourse-level exhaustivity implicature derived by reasoning with the QUD alternatives of the higher-level QUD.

5   General discussion

We started out with the question whether there is such a thing as discourse-level implicature, i.e. whether we can find pragmatic effects that require Gricean reasoning above the sentence level. We could show that some apparent discourse-level implicatures follow from the sum of sentence-level implicatures, so no reasoning with multi-sentence speech acts and no specific notion of discourse structure is required for their explanation, cf. the null theory. However, we have also found cases where the null theory predicts no discourse-level implicature, while Gricean reasoning at discourse level does, on the assumption that the discourse-level QUD serves as input to that reasoning process. Our experimental study focused on examples like (30), where the proposed QUD-based theory, but not the null theory, predicts the exhaustivity inference *Klaus didn't eat any pears*.

(30)   A: What (of apples, bananas, and pears) did Klaus eat?
         B: As for apples, he ate two. As for bananas, he ate four.

Exploiting the possibilities of contrastive topic marking by word order in German, our experiments show that in discourses like (30) B's answer indeed gives rise to the *Klaus didn't eat any pears* inference. That is, the implicature is drawn significantly more often in cases like (30) than under conditions where no exhaustivity implicature is expected. One such condition, tested in experiment 1, involves isolated sentences with a contrastive topic, such as *As for apples, he ate two*. Another case, studied in experiment 2, is clearly interrupted discourses ending with a conjunction followed by three dots (*and...*). Taken together experiment 1 and 2 provide clear evidence for discourse-level implicatures beyond the mere sum of sentence-level implicatures.

On the other hand we have seen that topicalization has a general effect reducing the number of exhaustivity inferences, even in cases like (30). That is, the answer in (30) is slightly but significantly "less exhaustive" than a direct answer without topicalization *Klaus ate two apples and four bananas*, as both experiments have shown. This suggests that discourse-level implicatures are drawn less frequently than sentence-level implicatures in comparable conditions. The question why this is the case remains open. However, one possible and rather straightforward explanation refers to differences in processing effort between drawing just sentence-level implicatures and both sentence- and discourse-level implicatures. The latter is expected to be more effortful and take more time, though showing this experimentally remains a task for the future.

Finally, coming back to the issue of contrastive topics, we mentioned in the introduction that most semantic theories of contrastive topic associate it, in one way or another, with strictly partial answerhood and the function of "cancelling exhaustivity" (Büring, 2003, to appear; Krifka, 1999; Hara and van Rooij, 2007). In this respect, contrastive topic is different from focus which is supposed to be exhaustive by default. This basic generalization was confirmed in our experiment 1. In cases like (31-B-b), where part of the object DP is marked as contrastive topic, the number of exhaustivity inferences ('Klaus did not eat any pears') was significantly lower (cf. 28,65%) than in (31-B-a), where the whole object DP is in focus (cf. 88.20% in table 2).

(31)   A: What (of apples, bananas and pears) did Klaus eat?
        B: a.   He ate [two apples]$_F$.
              b.   As for [apples]$_{CT}$ he ate [two]$_F$.

However, one of our main findings was that answers like (31-B-b) were also significantly less exhaustive than structures like (30-B) consisting of two sentences with a contrastive topic. None of the existing theories of contrastive topic has been able to account for this difference because until this day there has not been a theory of the base line case (30-B). Exhaustivity can be cancelled only if there is exhaustivity in the first place. In this paper we have formulated a theory that predicts an exhaustivity inference 'Klaus did not eat any pears' in (30-B). The standard semantics of contrastive topic accounts for the fact that (31-B-b) and each individual clause of (30-B) is non-exhaustive with respect to the question asked, whereas discourse-level implicature accounts for the exhaustivity of (30-B) as a whole.

6   Conclusion

In this paper we have argued for discourse-level implicatures beyond the sum of sentence-level implicatures. We have pointed out instances of quantity implicature pertaining to sequences of more than one sentence which cannot be captured by what we called the *null theory* of discourse-level implicature. We have formulated an alternative theory which exploits the hierarchical structure of discourse expressed in terms of questions under discussion. In that theory, discourse-level exhaustivity implicatures can result from reasoning with the alternatives provided by higher-level QUD dominating multiple sentences. In two experiments the QUD-based theory has been shown to make more accurate predictions than the null theory.

The particular complication introduced by quantity implicatures at discourse level is that they do not only require reasoning with multi-sentence speech acts, but also with discourse-level alternatives—the set of things the speaker could have said instead of the discourse he or she has actually produced. In single sentences, those alternative sets are traditionally derived from the focus structure of the sentence, by replacing the focused constituent by all possible expressions of the same semantic type. However, this approach is hard to apply to sequences of more than

one sentence which do not have a focus structure in the same sense as individual sentences do. The core of our proposal is that discourse-level alternatives used in the derivation of discourse-level quantity implicatures are constructed on the basis of the higher-level QUD. This provides additional motivation for the very concept of higher-level QUD and the hierarchical discourse structure based on questions under discussion. In other words, a QUD-based discourse structure is needed for an adequate account of discourse-level implicatures.

References

Asher, N., Lascarides, A., 2003. Logics of Conversation. Studies in Natural Language Processing. Cambridge University Press.

Bates, D., Maechler, M., Bolker, B., 2013. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-2. URL http://CRAN.R-project.org/package=lme4.

Büring, D., 2003. On D-trees, beans and B-accents. Linguistics and Philosophy 26, 511–545.

Büring, D., 2008. What's new (and what's given) in the theory of focus. In: Berson, S., Bratkievich, A., Bruhn, D., Amy Campbell, R. E., Giovine, A., Newbold, L., Marilola Perez, M. P.-B., Rhomieux, R. (Eds.), Proceedings of the Berkeley Linguistics Society Meeting. Berkeley Linguistics Society, Berkeley, CA, pp. 403–424.

Büring, D., to appear. (Contrastive) Topic. In: Féry, C., Ishihara, S. (Eds.), Handbook of Information Structure. Oxford University Press, Oxford.

Chemla, E., Spector, B., 2011. Experimental evidence for embedded scalar implicatures. Journal of Semantics 28 (3), 359–400.

De Kuthy, K., 2002. The information structure of discontinuous NPs in German. In: Proceedings of the 8th International HPSG Conference, Stanford. pp. 148–161.

Fanselow, G., Ćavar, D., 2002. Distributed deletion. In: Alexiadou, A. (Ed.), Theoretical approaches to universals. John Benjamins, pp. 65–107.

Fisseni, B., 2010. Focus: Interpretation? empirical investigations on focus interpretation. Ph.D. thesis, Universität Duisburg-Essen.

Fox, D., Katzir, R., 2011. On the characterization of alternatives. Natural Language Semantics 19, 87–107.

Geurts, B., 2007. Implicature as a discourse phenomenon. In: Puig-Waldmüller, E. (Ed.), Proceedings of Sinn und Bedeutung 11. Universitat Pompeu Fabra, Barcelona, Spain, pp. 261–275.

Geurts, B., 2010. Quantity Implicatures. Cambridge University Press.

Geurts, B., Pouscoulous, N., 2009. Embedded implicatures?!? Semantics and Pragmatics 2 (4), 1–34.

Grice, H. P., 1975. Logic and conversation. In: Cole, P., Morgan, J. L. (Eds.), Syntax and Semantics 3: Speech Acts. Academic Press, New York, pp. 41–58.

Groenendijk, J., Stokhof, M., 1984. Studies in the semantics of questions and the pragmatics of answers. Ph.D. thesis, University of Amsterdam.

Hamblin, C. L., 1973. Questions in Montague English. Foundations of Language 10, 41–53.

Hara, Y., van Rooij, R., 2007. Contrastive topics revisited: A simpler set of topic-alternatives, paper presented at NELS 38.

Keller, F., Gunasekharan, S., Mayo, N., Corley, M., 2009. Timing accuracy of web experiments: A case study using the webexp software package. Behavior Research Methods 41 (1), 1–12.

Krifka, M., 1999. Additive particles under stress. In: Proceedings of SALT 8. pp. 111–128.

Krifka, M., 2001. Quantifying into question acts. Natural Language Semantics 9, 1–40.

Krifka, M., 2008. Basic notions of information structure. Acta Linguistica Hungarica 55 (3), 243–276.

McCarthy, J., 1980. Circumscription—a form of non-monotonic reasoning. Artificial Intelligence 13, 27–39.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria, URL http://www.R-project.org/.

Roberts, C., 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. OSU Working Papers in Linguistics 49, 91–136.

Rooth, M. E., 1992. A theory of focus interpretation. Natural Language Semantics 1, 75–116.

Sauerland, U., 2004. Scalar implicatures in complex sentences. Linguistics and philosophy 27 (3), 367–391.

Schulz, K., van Rooij, R., 2006. Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. Linguistics and Philosophy 29 (2), 205–250.

Selkirk, E., 1995. Sentence prosody: Intonation, stress and phrasing. In: Goldsmith, J. (Ed.), Handbook of Phonological Theory. Blackwell, Oxford, UK, pp. 550–569.

Sperber, D., Wilson, D., 1986. Relevance. Communication and Cognition. Blackwell.

von Stechow, A., 1991. Focusing and backgrounding operators. In: Abraham, W. (Ed.), Discourse Particles. John Benjamins, pp. 37–84.

Zeevat, H., 2004. Contrastors. Journal of Semantics 21, 95–112.