Comparative modified numerals revisited: scalar implicatures, granularity and blindness to context[*]

Émile Enguehard
IJN, École Normale Supérieure, PSL

Abstract We present a theory of the pragmatics of comparative modified numerals such as "more than 5". Our proposal is based on the assumption that they trigger alternatives derived from discrete "granularity scales". This is sufficient to explain the basic pattern of scalar implicatures from comparative numerals; we then show how extra assumptions of blind exhaustification and QUD uncertainty refine the prediction in that we can explain how and when comparative numerals trigger irrelevance inferences.

1 Introduction

Comparative modified numerals, such as "more than five", are known to not trigger the scalar implicatures that bare numerals do. For instance, (1) is generally read to mean that John doesn't own more than three cars, which is often explained in the literature by a scalar implicature due to the presence of an alternative with "four":

(1)   John owns three cars.

Instead, unembedded comparative numerals give rise to so-called ignorance inferences (the speaker doesn't know the number) or to irrelevance inferences (the question under discussion makes the precise number irrelevant). This is seen in (2):

(2)   John owns more than three cars.
         ↛ John doesn't own more than four cars.

Cummins, Sauerland & Solt (2012) note that comparative modified numerals may in fact receive enriched readings when they combine with "round" numerals, as in (3).

(3)   There are more than 90 people in this room.
         ⇝ There aren't more than 100.

We can extend this observation to non-round numbers. Specifically, enriched readings of non-round comparative numerals are sometimes available when they refer to continuous quantities, as in (4).

(4)   a. John walked more than 7 kilometers to get home.
          ⤳ He didn't walk 8 kilometers.

       b. John has been working there for more than 22 years.
          ⤳ He hasn't been there for 23 years.

2   Comparative numerals in the literature

The most straightforward approach towards comparative numerals, exemplified by Mayr 2013, takes them to evoke alternatives such as "exactly *n*" or "more than *n*" for integer *n*. Those alternatives are used in an exhaustification mechanism that evaluates them in context, resulting in the prediction that (2) triggers the *ignorance inference* that the speaker thinks it's possible that John read only four books, or that he read more.

Another approach is that of Fox & Hackl (2006); they introduce the *Universal Density of Measurement* (UDM) hypothesis, according to which the alternatives of modified numerals are based on all real numbers. Together with the assumption that the mechanism giving rise to scalar implicatures is blind to contextual information, they predict that comparative numerals cannot receive enriched meanings, not even ignorance implicatures.

Neither approach provides an immediate explanation for the full pattern we outlined above. The theory I am going to present is based on discrete scales; in section 6.3, we'll see how equivalent predictions could be derived from dense scales.

3   First proposal

3.1   Formal exposition

To capture the difference between (2) and (3), we need to make reference to a notion of *granularity* in our definition of alternatives. This is provided by Cummins et al.'s (2012) *granularity scales*. A granularity scale is the scale of all numerals of a certain "roundness" such as:

   a. $\langle 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, \ldots \rangle$

   b. $\langle 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, \ldots \rangle$

   c. $\langle 0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, \ldots \rangle$
      etc.

We assume that every modified numeral is interpreted within a granularity scale, and this determines its formal alternatives. Formally, if n is interpreted with the granularity scale S:

(A)   alt("more than n") = {"m", "more than m", "exactly m"|$m \in S$}

where the semantics of "n" are assumed to be "n or more". The granularity scale chosen is the coarsest one available for the numeral (the units for "23", tens for "20", etc.).[1] Complete utterances also have formal alternatives which are obtained through replacing any of their constituents by their own formal alternatives.

We assume that any utterance answers a tacit "Question Under Discussion" (QUD), and for a given QUD Q there is a fixed set of "relevant" propositions R_Q. Four our purposes, we will consider two kinds of question:

1. "How many?" questions like "How many cars does John own?". We will call these *total questions*, and assume they make all alternatives we discuss relevant.

2. *Partial questions* like "Does John own more than three cars?". We will assume these questions make all other answers than (2) and its negation irrelevant.

Finally, we adopt the following fairly standard treatment of implicatures, except that we treat them as obligatory:

1. If ϕ answers Q, for all $\psi \in \text{alt}(\phi) \cap R_Q$ such that ϕ doesn't entail ψ, the hearer understands that the speaker doesn't believe that ψ is true (primary implicature). The set of such propositions is $\text{alt}'(\phi; Q)$.

2. If $\psi \in \text{alt}'(\phi; Q)$ is innocently excludable in the sense of Fox 2007,[2] the hearer may infer that the speaker believes that ψ is false (secondary implicature).

In a given context, there may be several salient questions to which ϕ is a relevant answer. Upon hearing ϕ, the hearer then gets a disjunctive inference: for at least one

---

1 Possibly the set of available granularity scales is contextually or culturally dependent, though there is no clear evidence for or against it. Possibly too, context may sometimes select a more precise granularity scale; see section 6. Finally, in some cases the "coarsest scale available" might be poorly defined, as it seems that non-linear scales like $\langle 0, 1, 10, 100, \ldots \rangle$ are available. (Cummins et al. 2012)

2 ψ is innocently excludable within $\text{alt}'(\phi; Q)$ if and only if every maximal co-deniable subset of $\text{alt}'(\phi; Q)$ contains ψ. A subset is co-deniable if and only if the conjunction of the negations of all its elements isn't a contradiction. Intuitively, this means that ψ is innocently excludable if it can be taken to be false together with any other alternatives without resulting in a contradiction.

of the salient questions, the primary and secondary implicatures as computed above hold, and the alternatives not in R_Q are irrelevant to the speaker's intention. Certain choices of Q may be eliminated if they result in pragmatically odd meanings.

3.2 Applying the mechanism

We will now look at how the theory above applies to (2) and (3). Per (*A*), the alternatives to (3) are those given in (5). Those that are neither weaker than nor incompatible with (3) are all those that involve numerals above 100, as in (6). Among those there is a single weakest element (7), involving a bare numeral.

(5)   There are (more than / exactly / ∅) (.../80/90/100/...) people in this room.

(6)   There are (more than / exactly / ∅) (100/110/120/...) people in this room.

(7)   There are 100 people in this room.

If the most salient QUD is the total one, which makes all alternatives relevant, step 1 and step 2 above result in the listener's inference that the alternatives in (6) are false, as it can be verified that they are all innocently excludable. This is equivalent to just inferring that the weakest one, (7), is false. Thus we get the enriched meaning: there are more than 90 people, but fewer than 100.

If the most salient QUD is the partial QUD "are there more than 90 people here?", all alternatives other than (3) are irrelevant. We predict (3) to not receive an enrichment.

If there is uncertainty about the QUD, we predict a disjunctive inference: either (7) is false, or all the propositions in (6) are irrelevant.

Those predictions are consistent with the fact that (3) is acceptable when there are clearly much more than 90 people only if the partial QUD is salient, as in (8a), while it is quite degraded under an explicit total QUD as in (8b).

(8)   Context: it is visible that there are around 200-300 people.

     a. A: We've got 90 chairs.
       B: There are more than 90 people here.

     b. A: How many people are there?
       #B: There are more than 90 people here.

The treatment of (3) extends to any usage of round numerals or any description of dense quantities. For instance, the case of (4a) would follow the same pattern, only using the unit scale instead of the tens. The alternatives of (4a) are listed in (9), with the ones that are non-weaker and co-tenable with (4a) in (10). This set has exactly the same structure as (6); there is a weakest element (11), counterpart to (7).

(9)   John walked (more than / exactly / ∅) (.../6/7/8/...) kilometres to get home.

(10)   John walked (more than / exactly / ∅) (8/9/10/...) kilometres to get home.

(11)   John walked 8 kilometres to get home.

Without repeating the detail of it, we predict for (4a) a disjunctive inference completely parallel to the one we predict for (3): either all the alternatives in (10) are irrelevant, or (11) is false.

The other case we want to account for is that of non-round numerals being used to describe discrete quantities, as in (2). The alternatives to (2) given by (A) are in direct correspondence to those of the previous example (3); they are listed in (12). What changes is that we no longer have a single minimal non-weaker alternative: the bare numeral counterpart to (7), given in (13), is not non-weaker than the initial proposition (2): if John owns more than three cars, he owns four cars. Instead, among those alternatives that are non-weaker than and not incompatible with (2), there are two minimal elements, (14a) and (14b). They are minimal in the sense that there are no elements in the set that are weaker than either.

(12)   John owns (exactly / more than / ∅) (1 / 2 / 3 / 4 / ...) car(s).

(13)   John owns four cars.

(14)   a. John owns exactly four cars.

        b. John owns more than four cars.

(14a) and (14b) are not innocently excludable: given (2), negating either one entails that the other is true. This is the well-known pattern of symmetric alternatives. Hence, no secondary implicature should arise from them: if the salient QUD is the total question, which makes all alternatives relevant, we predict an *ignorance inference*. The speaker doesn't know whether John has exactly four or more than four cars. This is also what Mayr (2013) predicts, and for the same reasons.

As before, if the salient QUD is the partial question, we predict a non-enriched reading. If both QUD are salient, we predict a disjunctive inference: either the speaker is ignorant about (14a) and (14b), or they are irrelevant.

To conclude this section, we can note that the key property that distinguishes (3) and (4a) from (2) is whether there are possible worlds where the quantity being discussed is both above the stated bound and below the next item on the granularity scale. For instance, it is possible to walk more than 7 and less than 8 kilometres, or for there to be more than 90 and less than 100 people, but it is not possible to own more than 3 and less than 4 cars. This is why we predict a secondary scalar implicature for (4a) and (3), but not for (2). The fact that (3) is about a dense quantity and (4a) is about a discrete one has no consequence in our theory.

4   Refining the proposal: structural entailment and obligatory irrelevance inferences

In the discussion so far, we have adopted the contextual definition of entailment: ϕ (contextually) entails ψ if all contextually likely worlds where ϕ is true are also worlds where ψ is true. However, certain theories, including the UDM treatment of modified numerals by Fox & Hackl (2006), make use of a stronger notion of entailment, *structural* (or *logical*) entailment. ϕ structurally entails ψ if it is true at a smaller set of worlds, i.e. if all worlds that make ϕ true also make ψ true, independently of the current context. In the rest of this section, I will argue for structural entailment by showing how it modifies the predictions of the proposal, and why such changes are for the better.

Recall that propositions containing "more than n" and involving round numerals or dense quantities have a single minimally stronger alternative, containing "$n+1$ (or more)." When non-round numerals are used to describe discrete quantities, the corresponding alternative is equivalent to the original proposition in the sense of contextual entailment. For instance, (15) (repeated from (2)) has the alternatives in (16) (ignoring strictly weaker and incompatible ones). The weakest among them is (17), which is contextually equivalent to (15). Because it is equivalent to the original proposition (and therefore weaker in the general sense), it isn't considered by the procedure described in section 3.1.

(15)   John owns more than three cars.

(16)   John owns (more than / exactly / ∅) (4/5/6...) cars.

(17)   John owns four cars.

With a structural definition of entailment, it is no longer true that (17) is equivalent to (15), as long as we take the fact that cars come in integer quantities to be peculiar to the context.[3]

1. Our prediction under the partial QUD isn't affected by this fact, since all alternatives are irrelevant anyway.

2. Under the total QUD, all alternatives are relevant. They are also all innocently excludable, and denying them is equivalent to denying the single minimal one, namely (17). Thus, we predict the following enriched meaning:

---

3 There are plausible worlds where John owns a fraction of a car, say if he shares it with a relative. I do not think we should rely on this fact, though: first, it doesn't intuitively spring to mind; second, it may not be true in a given time and place, say for legal reasons, and yet this shouldn't affect the linguistic facts; third, all what we say about (2) also applies to "Hexagons have more than 5 sides" as far as I can tell; yet, it is hard to imagine what a shape having a fraction of a side would correspond to.

John owns more than three cars, but fewer than four. This is a contextual contradiction: the sentence should be infelicitous.[4]

3. When there is uncertainty about the QUD, we ought to predict again a disjunctive inference: either John owns fewer than 4 cars, or the number of cars he owns is irrelevant. Because the first disjunct is impossible, the listener has to accept the second one. Hence, what we actually predict is an obligatory *irrelevance inference*: the listener concludes that the speaker is interested in the partial QUD, even if it weren't very salient.

Compared to the contextual definition of entailment, we no longer predict the possibility of ignorance inferences. This is at odds with some of the literature, such as Mayr 2013, where such ignorance inferences are taken to be fact, and the theory is designed to derive them. Buccola & Haida (2017) are among some that argue that they don't exist. Indeed, in dialogues where the total QUD is made very explicit, such as (4), while "at least" numerals are felicitous even with non-round numbers, "more than" numerals are somewhat degraded. Given that "at least" numerals are well-known to trigger ignorance inferences (Schwarz 2016), this suggests that "more than" numerals are different. If, as we predict, they trigger obligatory selection of the partial QUD, their infelicity as an answer to the total one is readily explained.

(18)   A: How many dogs does Ann own? Three? Seven? Twelve? I have no idea.

a. B: At least 5 / 10 / 13.

b. B: More than ??5 / 10 / ??13.   (adapted from Buccola & Haida 2017)

Another case where comparative numerals are degraded is when other partial QUDs, involving other thresholds, are very salient in the context. For instance, in the context of (19), 18 (or 21) is a much more relevant threshold than 33, making the accommodation of the partial QUD centred on 33 difficult. The superlative numeral is felicitous as it is a valid answer to the total QUD, which is, presumably, always relevant.

(19)   (We're in a bar.)   (Benjamin Spector (p.c.))
        Of course she can drink, she's (at least / #more than) 33.

Given such examples, the choice of structural entailments, without changing our architecture, seems to provide better empirical adequacy.

To summarize the pattern that our theory predicts under both definitions of entailment:

---

4 Note that even the primary implicature is already a contextual contradiction: John owns more than three cars, but the speaker isn't sure he owns four. Our prediction doesn't depend on secondary implicatures being obligatory.

1. With both definitions, we predict two readings when the quantity in question could be measured on a more precise scale as in (3), either because round numerals are used or because the quantity is dense. There is first an enriched reading with a scalar implicature, associated to a total QUD ("how many?"), and second, a non-enriched reading associated to a partial QUD. This seems empirically valid.

2. With the contextual definition, when the quantity cannot be measured any more precisely, because non-round numerals are used about discrete quantities, we predict no secondary implicature, plus an ignorance inference only when the total QUD is salient.

3. In the structural variant, when the quantity cannot be measured any more precisely, we predict no secondary implicature, and an obligatory irrelevance inference: the listener has to accommodate the partial QUD regardless of context. Evidence points in favour of this prediction over the one above.[5]

5 Embedded environments

5.1 Universal quantification

In this section, we will explore the predictions of our proposal when universal quantifiers are involved, and find further evidence teasing apart contextual and structural entailment.

To begin with, we will look at the case of discrete quantities and non-round numerals as in (20). Its alternatives (omitting strictly weaker and incompatible ones) are listed in (21). This set includes a logically minimal element, (22). The rest of the set has two incomparable minimal elements, (23a) and (23b).

(20)   Every student of mine made more than 7 mistakes.

(21)   Every student of mine made (more than / exactly / ∅) (8/9/10/...) mistakes.

(22)   Every student of mine made 8 mistakes.

(23)   a. Every student of mine made more than 8 mistakes.

        b. Every student of mine made exactly 8 mistakes.

Assume a QUD that makes all alternatives relevant is salient. Under the contextual definition of entailment, (22) is equivalent to (20), and therefore weaker in a general sense. We only derive implicatures from non-weaker alternatives, so (22) is ignored. All other alternatives are innocently excludable, and we predict them to

---

5 Note that while the structural theory doesn't allow for an ignorance inference, the reading it predicts is *compatible* with speaker ignorance.

trigger secondary implicatures. Because denying them all is equivalent to denying the two minimal ones, (23a) and (23b), the listener's final inference upon hearing (20) is that not all students made exactly 8 mistakes, nor did they all do more than 8 mistakes. In other words, the enriched meaning of (20) is that some students did 8 mistakes exactly, and some did more.

Under the structural definition of entailment, (22) is non-weaker than (20). Again, all alternatives in (21) are innocently excludable, and denying them all is tantamount to denying (22). Hence, we predict an inference that some students did fewer than 8 mistakes. This is a contextual contradiction, and (20) ought to be infelicitous. However, if a partial QUD is at least somewhat salient in the context, we predict it to be obligatorily selected (irrelevance inference).

Thus, there is no difference with the unembedded case as far as the structural definition of entailment is concerned. For the contextual definition, we get a dual secondary implicature instead of an ignorance inference.

Again, we can try to find out whether this dual implicature is possible by comparing to "at least $n+1$", for which this dual implicature is well-known to be there. If they triggered the same inferences, we would expect B's utterance in (24b) to be an equally valid continuation as B's utterance in (24a). Though judgements aren't clear, there appears to be a contrast. Such data point towards the structural definition of entailment as the better one within our proposal.

(24)   Context: you need to make fewer than 10 mistakes to pass.

      a. A: Every student of mine made at least 7 mistakes.
         B: Well, at least some of them passed.

      b. A: Every student of mine made more than 7 mistakes.
         ?B: Well, at least some of them passed.

In the case of dense quantities and round numerals, the definition of entailment makes no difference. (25) has the alternatives in (26), amongst which there is a single minimal element (27).

(25)   There were more than 80 people at every class.

(26)   There were (more than / exactly / $\varnothing$) (90/100/...) people at every class.

(27)   There were 90 people at every class.

If all alternatives in (26) are relevant, since they are all innocently excludable, we may derive secondary implicatures from all of them. The inference this results in is the negation of (27): hence the enriched meaning we predict is that at certain classes there were fewer than 90 people.[6]

---

6 The stronger inference that there were fewer than 90 people at every class is attractive, but it is easy

5.2 Negation

We now describe the predictions of the mechanism when comparative numerals are embedded under negation.

We first have to specify what the alternatives of a sentence involving "not more than n." We will assume that if n is interpreted within the granularity scale S, then this phrase has the alternatives in (A').

(A') alt("not more than n") =

$$\{\text{"fewer than } m\text{"}, \text{"not more than } m\text{"}, \text{"exactly } m\text{"} \,/\, m \in S\}$$

The first two series of elements are negated forms of alternatives of "more than n", assuming that negating a bare numeral produces "fewer"; the third one is an alternative of "more than n".

We need to assume this is the alternative set to not make undesirable predictions: if "more than m" and "m (or more)" are included, as we would expect if deleting negation was a possible way of deriving alternatives, we run into the symmetry problem: there are no innocently excludable alternatives at all, and we predict ignorance inferences across the board; we will see how this doesn't matches the data. If "exactly m" is not included, we predict a systematic enrichment of "not more than n" into meaning "exactly m"; we will discuss below whether this is really wrong.[7]

Let's first look at an example using a round numeral, (28). Its not strictly weaker nor incompatible alternatives, as derived from (A'), are listed in (29). Among them there are two minimal elements, (30a) and (30b). Among the rest, there is a single minimal element, (30c).

(28)   There aren't more than 80 people here.

(29)   a. There aren't more than .../60/70 people here.

---

to cancel, thus it cannot derive from the obligatory mechanism we're studying here:

(i)   How many people came to the class?

   a. There were more than 80 people at every class, and once they were 127!

   b. # There were more than 80 people at every class, in fact the lowest number was 127.

7 A tentative justification for this necessary assumption is that deleting negation is allowed, but some alternatives are excluded because expressing lower bounds or upper bounds has an evaluative component. Specifically, sentences using "more than" are used when the quantity is perceived to be high, etc. This would explain why "more than" and "fewer than" do not compete.

We could avoid the assumption altogether if we assumed than the exhaustification procedure only makes reference to stronger alternatives (as opposed to non-weaker ones). Then "more than" and "fewer than" would not compete because there is no entailment relationship between them.

      b. There are fewer than .../60/70/80 people here.

      c. There are exactly .../60/70/80 people here.

(30)   a. There are fewer than 80 people here.

      b. There are exactly 80 people here.

      c. There aren't more than 70 people here.

As usual, under a partial QUD, all those alternatives are irrelevant and we predict no enrichment.

Given (28), negating (30a) entails that (30b) is true and vice-versa. This means that they are not innocently excludable. Hence, under a total QUD, we predict the listener to draw an ignorance inference about (30a) and (30b). However, (30c) and all stronger alternatives are innocently excludable and may trigger secondary implicatures. The full enrichment that we predict for (28) is therefore that there are between 70 and 80 people, and the speaker considers it possible that there are as many as 80, but isn't sure.

This is, as far as I can tell, empirically valid: B's answer in (31) is degraded because it violates the inference that there are more than 100 people,[8] and (32) is degraded when ignorance is difficult to accommodate.

(31)   Context: there are clearly fewer than 50 people.
A: How many people are there here?
#B: There aren't more than 200.

(32)   Context: I have 47.54€ on my account.

      a. ? I just checked my balance, and I haven't got more than 50€.

      b. I haven't checked my balance, but I haven't got more than 50€.

For the case of discrete quantities and non-round numerals, we will work off (33) as an example. Its alternatives (in (34)) have the same structure as those of (28): there are two minimal ones, (35a) and (35b), and the rest of them have a minimal element, (35c).

(33)   John doesn't own more than 3 cars.

(34)   a. John doesn't own more than 1/2 cars.

      b. John owns fewer than 1/2/3 cars.

      c. John owns exactly 1/2/3 cars.

(35)   a. John owns fewer than 3 cars.

---

8 The primary implicature (B isn't sure there are fewer than 100 people) is in fact enough to explain why B's answer is degraded. Note that there is no ignorance inference about whether there are 100 people: B's answer is felicitous when there are clearly more than 100 people.

b. John owns exactly 3 cars.

c. John doesn't own more than 2 cars.

If entailment is defined contextually, (35c) is equivalent to (35a), and there are no innocently excludable alternatives. We predict (33) to be felicitous under the total QUD, and to trigger an ignorance inference.

If entailment is defined structurally, the computation starts out the same way as for (28). Under the total QUD, all alternatives are relevant. (35b) and (35a) are symmetric alternatives, and therefore not innocently excludable; we predict an ignorance inference about them. (35c) and all stronger alternatives are innocently excludable. The secondary implicature derived from (35c) leads to a contextual contradiction: John owns more than 2 cars and not more than 3, but the speaker isn't sure he owns exactly 3. On the other hand, if (35c) only results in a primary implicature, there is no contradiction. Thus, the prediction depends on whether we assume secondary implicatures are obligatory: if they are, we predict infelicity from the contextual contradiction; if they aren't, we predict felicity under the total QUD, together with an ignorance inference.

There is no evidence that "not more than *n*" is infelicitous under the total QUD; for instance, (36) is fine. Given that this is compatible with both definitions of entailment, this doesn't help us towards choosing between them.

(36)   A: How many people will come? Five?
         B: There won't be more than three of them.

6 Issues

6.1 Further issues regarding negated comparative numerals

In this section, we describe various interesting issues that our proposal doesn't account for as well as might be hoped.

The first one is that the behaviour of negated comparative numerals is more complex than what we describe in section 5.2. To begin with, there are in fact two ways to negate comparative modified numerals in English, exemplified by (37) (repeated from (33)) and (38).

(37)   John doesn't own more than 3 cars.

(38)   John owns no more than 3 cars.

There is no reason why we would predict any difference between these two constructs. Yet, they are not completely equivalent. For instance, (37) is a fine answer to the partial QUD, in which case its meaning isn't enriched, as we predict. (38), though, is somewhat degraded as an answer to the explicit question "does John

own more than 3 cars?," and it isn't clear what its meaning is. Similar facts are already described by Nouwen (2008), who notes that the "not" construct, where the negation has a wider surface scope and perhaps wider scope at LF too, is presumably more appropriate for "denial" — what we call answering the partial QUD here is probably a superset of denial.[9]

As an answer to the total QUD, "no more" can trigger the inferences that we predict: the listener draws an ignorance inference about whether the bound is reached for both discrete and dense quantities, as in (40) and (39), as well as a scalar implicature in the dense case, as in (40).

(39)   John owns no more than 3 cars.
⤳ The speaker isn't sure John has as many as 3 cars, but they think it's possible. (possible reading)

(40)   John drank no more than 2 litres of beer in his life.
⤳ John drank more than one litre, and the speaker isn't sure whether he reached two. (possible reading)

However, "no more" (and perhaps "not more" too) has an additional, stronger reading with no ignorance or irrelevance inference, which is the focus of Nouwen's (2008) analysis. This reading is one where the speaker is in fact fully answering the total QUD; for instance, (41) has a reading where the speaker believes that John has exactly 7 books, as shown by the validity of the continuation. This reading comes with a strong evaluative component: we understand that the speaker believes that the quantity in question is small.[10]

(41)   John has no more than 7 books in his library.
(I counted them all.)

We would predict this reading if we didn't have the "exactly" alternative in (A'), which in turn would be the case if we assumed deleting negation in alternatives is not licit. Then, (35a) would be the single minimal alternative to (33). A possible explanation for the difference between "no more" and "not more" would then be that the negation in "no more" is harder to delete than the sentential one; this merits further exploration. Alternatively, it could be that the "exactly" alternative is never present: then, all the examples we gave of readings where the speaker is ignorant would just be instances of primary implicatures not being strengthened, rather than obligatory ignorance inferences: it remains to be seen how those two predictions can

---

9 Another reason for this pattern could be that stressing the negation is a way to make the partial QUD more salient, and it is not possible in (38).

10 In fact most examples in this paper carry a suggestion that the quantity is small when stating an upper bound and high when stating a lower bound, at least when they're interpreted under the total QUD.

be told apart.[11] In any case, since the strong reading seems to be accessible only in a restricted set of contexts, those where the quantity is perceived to be very small, while the weaker reading where the speaker is ignorant is more generally accessible, I chose to select the alternatives set that leads to generating the weaker reading.

6.2   The case of necessity modals

In this section, we review a puzzle on the behaviour of comparative modified numerals embedded under deontic necessity modals. Because they have the same structure, we predict necessity modals to behave similarly to universal quantifiers like "every", that is:

1. When using round numerals or talking about dense quantities, there is a secondary scalar implicature.

2. When using non-round numerals and talking about discrete quantities, there is an obligatory irrelevance inference.

This is indeed what we observe with epistemic modals. (42) triggers an irrelevance inference, and (43) a scalar implicature.

(42)   (For all I know,) John must have more than 3 cars.
⤳ It is irrelevant how many he has.

(43)   There must be more than 90 people here.
⤳ There may be fewer than 100.

However, some deontic examples appear to have readings that we do not predict. (44), for instance, should trigger an irrelevance inference: its minimal alternative is (45), and strengthening that would result in the nonsensical proposition that John may answer strictly less than 9 questions and pass. Instead, (44) is a felicitous description of a rule stating that John may pass by answering 9 questions or by answering more.

(44)   John must answer more than 8 questions to pass.

(45)   John must answer 9 questions to pass.

(46)   a. John must answer 9 questions exactly to pass.

b. John must answer more than 9 questions to pass.

---

11 Yet another explanation could be that the "exactly" alternative may *sometimes* not be available, though this is harder to explain; certainly it cannot be irrelevant while the "fewer than" one is. Possibly, when prior expectations about the quantity are that it should be much higher, the "exactly" alternative is overly informative as the "no more" one already locates us in a sufficiently small corner of the epistemic space; this can be formalised within existing game-theoretical models of pragmatics.

The actual reading of (44) is what our proposal would straightforwardly predict if we had kept to contextual entailment. Indeed, (45) is contextually equivalent to (44). This makes the alternatives in (46) the minimal ones; as they are innocently excludable, the final prediction is that (44) triggers the inference that they do not hold.

This fact puts a dent in the otherwise strong empirical validity of the structural variant. It is possible to rescue it, but it requires some questionable assumptions on modality. The idea is that the problematic inference obtained by denying (45) is not as untenable as other problematic inferences we had before. If deontic modality is assumed to be partly blind to world knowledge, and in particular, if the fact that it is only possible to answer integer numbers of questions is not part of the deontic modal base, then denying (45) isn't contradictory even together with (44): (44) asserts that answering between 8 and 9 questions is compatible *with the rules*, regardless of whether it is feasible or not. It is in fact possible to express explicitly such nonsensical rules, as in (47):

(47) Context: There are 43 voters in a yes/no ballot.
Per the rules, the change will be rejected if there are at least 21.5 "no" votes.

Another problem arises when deontic modals embed comparative numerals describing dense quantities or using round numbers. For instance, (48) is predicted to trigger the inference that it isn't necessary to run 4km to get a prize. It actually has a stronger reading where running any distance above 3km is enough to get a prize.

(48) To get a prize, you must run more than 3km.

This strong reading is one we could generate most straightforwardly from dense scales, following Fox & Hackl's (2006) UDM hypothesis. Arguably, our proposal also generates it. Indeed, in deontic contexts, it is easy to select precise granularity scales even where coarse ones are available: this is shown by the fact that (49), the counterpart to (44) using the round number 10, doesn't necessarily get the reading that answering fewer than 20 questions can be sufficient: instead it can mean that answering 11 questions can be sufficient.

(49) John must answer more than 10 questions to pass.

While we have no explanation for this fact, we take it to imply that (48) can also be interpreted with a very precise granularity scale, such as $\langle 3, 3.001, 3.002, \ldots \rangle$. Then, we predict it to mean something like "running 3.001km is enough to get a prize", which is mathematically different from the reading we're looking for, but indistinguishable from it in practice.

6.3 Alternative theories

The theory we've discussed so far is based on stipulated discrete alternative scales, of the kind that are traditionally used in the literature. In this section, we attempt to sketch alternative theories that would at least capture the pattern in section 1, while being based on a single dense scale, as per Fox & Hackl's (2006) UDM hypothesis.

A first approach would be to assert that granularity depends on the QUD. Thus, instead of a single "total QUD", we would have a variety of them for different granularities. Upon hearing "more than 10" when talking about discrete quantities, listeners would infer that the QUD is either the partial one, or the total one with granularity ten, because the utterance wouldn't be a relevant answer to a coarser total QUD, and would be incoherent as an answer to a more precise one. The predictions would be the same as what we've reviewed here, with an additional one: using, say "more than 10" should trigger irrelevance inferences about alternatives involving 11, 12, etc., which seems like a good result. I haven't presented this theory as it wasn't my initial approach, and its formalization could perhaps present some challenges.

We can try to stay closer to Fox & Hackl's (2006) analysis by relying on closure conditions on the alternatives set to derive the lack of scalar implicature, in the spirit of Fox & Katzir 2011 (Danny Fox and Benjamin Spector, p.c.). Assume that alternatives derived from the entire dense scale are part of the basic alternative set, before relevance considerations come into play. Assume also the following closure condition: alternatives that are contextually equivalent to the utterance have to be relevant. When describing a dense quantity, "more than 7 kms" and "more than 7.5 kms" are not contextually equivalent: we predict it should be possible for the latter alternative to be pruned, keeping only the "8 (or more)" alternative and deriving a secondary scalar implicature from it. On the other hand, when talking about discrete quantities, "more than 3.5 cars" and "4 (or more) cars" are contextually equivalent to "more than 3 cars": therefore they cannot be pruned, and exhaustification cannot take place.

While this theory predicts the lack of scalar implicatures for (2) (repeated below as (50)) and their presence for (4a) (repeated as (51)), it doesn't derive the obligatory irrelevance inference, and it doesn't explain why round numerals (as in (3), repeated as (52)) behave like dense quantities. Indeed, recall that our proposal makes no difference between (51) and (52) because the crucial property for us is whether it is tenable to deny the next bare numeral alternative: John may walk more than 7 and fewer than 8kms, there may be more than 90 and fewer than 100 people, and John may not own more than 3 and fewer than 4 cars. In the alternative proposal we just sketched, the crucial property is whether some alternatives are contextually equivalent to the utterance: John walking more than 7.1kms isn't equivalent to John walking more than 7kms, but John owning more than 3.1 cars is equivalent to John

owning more than 3 cars, and there being more than 90.1 people is equivalent to there being more than 90 people. For the same reason that it isn't licit to rule irrelevant whether John owns more than 3.1 cars, it shouldn't be licit to rule irrelevant whether there are more than 90.1 people. This groups (52) together with (50), which is wrong.

(50)   John owns more than 3 cars.

(51)   John walked more than 7 kilometres to get home.

(52)   There are more than 90 people here.

In the hypothetical theories we just sketched, the possibility of interpreting within the full dense scale should help account for problematic cases involving deontic modals such as (48), though one would need to explain why epistemic cases like (43) behave differently. The case of (44) is problematic for all the theories we discussed.

7   Discussion and perspectives

As we've seen, the combination of a structural view of entailment and the introduction of granularity scales into theories of alternatives lets us provide fairly comprehensive empirical coverage of comparative numerals.[12] We can explain:

a. Why they don't trigger ignorance implicatures, in contrast with so-called superlative numerals.

b. Why and when they trigger irrelevance implicatures.

c. Why and when they trigger standard scalar implicatures (contrary to previous assumptions that they never did).

Straightforward theories based on discrete alternative scales (including our initial "contextual" proposal) fail to explain the first point, while in theories based on dense scales, as per Fox & Hackl's (2006) UDM hypothesis, it isn't straightforward to explain why scalar implicatures occur in cases involving discrete quantities. Our proposal doesn't directly bases its predictions on whether the quantity being discussed is dense.

One of the important ingredients of this proposal, or of its hypothetical alternative formulations, is mandatory blind exhaustification. Mandatory blind exhaustification has been proposed for other phenomena as well (cf. for instance Magri (2009)), but it runs against the common idea that implicatures are cancellable. The architecture

---

12 I haven't given any example involving a "fewer than" numerals, but if their alternatives are those in (A'), the predictions we make for them are completely symmetric to those we make for "more than" numerals, which seems correct.

of our model, where different exhaustifications are possible depending on the QUD, may perhaps be used to explain this fact in more general account of implicatures, where we'll try to characterize cancellable and "pragmatic" inferences as deriving from QUD selection. It remains to be seen how far this can be pushed.

As far as modified numerals are concerned, the natural next step from here is to address the behaviour of superlative numerals, and first, assess their interaction with granularity. Besides, an important trait of modified numerals that has only been briefly touched on in this paper is the presuppositions or evaluative component on prior expectations and desires that they carry. This is particularly true for expressions like "only *n*" or, as we mentioned, "no more than *n*" and (even more so) "no less/fewer than *n*", in their "exactly" readings. The formalism we have been using here doesn't seem to be able to account for such effects: it would be desirable to find out how to solve this, and, to begin with, to fully examine the interaction between the sort of implicatures we're interested in and a-priori expectations.

References

Buccola, Brian & Andreas Haida. 2017. Obligatory irrelevance and the computation of ignorance inferences. Ms. (July 31, 2017); lingbuzz/003600.

Cummins, Chris, Uli Sauerland & Stephanie Solt. 2012. Granularity and scalar implicature in numerical expressions. *Linguistics and Philosophy* 35(2). 135–169. doi:10.1007/s10988-012-9114-0.

Fox, Danny. 2007. Free choice disjunction and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*, 71–120. London: Palgrave Macmillan. doi:10.1057/9780230210752_4.

Fox, Danny & Martin Hackl. 2006. The universal density of measurement. *Linguistics and Philosophy* 29(5). 537–586. doi:10.1007/s10988-006-9004-4.

Fox, Danny & Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19(1). 87–107. doi:10.1007/s11050-010-9065-3.

Magri, Giorgio. 2009. A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics* 17(3). 245–297. doi:10.1007/s11050-009-9042-x.

Mayr, Clemens. 2013. Implicatures of modified numerals. In Ivano Caponigro & Carlo Cecchetto (eds.), *From Grammar to Meaning: the Spontaneous Logicality of Language*, 139–171. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139519328.009.

Nouwen, Rick. 2008. Upper-bounded no more: the exhaustive interpretation of non-strict comparison. *Natural Language Semantics* 16(4). 271–295. doi:10.1007/s11050-008-9034-2.

Schwarz, Bernhard. 2016. Consistency preservation in quantity implicature: The case of *at least*. *Semantics and Pragmatics* 9(1). 1–47. doi:10.3765/sp.9.1.

Émile Enguehard
Institut Jean Nicod
École Normale Supérieure
29 rue d'Ulm
75005 Paris
France
emile.enguehard@ens.fr