Can Scalar Implicatures be drawn from Modified Numerals?

On the processing profile of modified numerals and their potential scalar implicatures under universal quantifiers

Master's thesis of Erlinde Meertens

Linguistics: The Study of the Language Faculty

Supervisors: Rick Nouwen and Yaron McNabb

Second reader: Bert le Bruyn

Utrecht institute of Linguistics

June 2015

Universiteit Utrecht

Acknowledgements

The first person I would like to thank is my first supervisor Rick. In my mind, the best supervisor is always critical and sharp, but in an encouraging and constructive way. During the writing of this thesis, Rick did not only tick these boxes. I also realized slowly how knowledgeable and passionate he is, which inspired me a lot.

Of course, I also like to thank Yaron for his supervision and his involvedness in this thesis. Thanks especially for the fruitful statistiscs sessions. For the first time ever, I felt like I understood the tests I was running. Also, thanks for the constructive feedback and the encouraging words I sometimes needed.

I want to thank Bert le Bruyn for his willingness to read my thesis in the couple of days I gave him for it.

Further, words of thanks need to be attributed to some of the people working at the UiL-OTS. First, the technical staff and in particular Chris van Run, who helped me programming the experiment and dedicated many hours of his valuable time to my experiment, of which I am really thankful. Without him I would not have been able to test my participants. Also, thanks to Iris Mulders who gave me access to her participant database. This allowed me to test 10 extra people, a big difference in the reliability of the experiment. Speaking of participants, this thesis could not have been written without the voluntary participants, who sometimes came from impossible places at impossible times.

A special shout goes to Jan Jongboom, who left Firefox OS for what it was to be my ZEP programming assistant for one day. Without him, I would not have been able to run the experiment.

I want to thank my brothers, sister and the other great people, including Barbara, Evelien, Jeroen, Mark, Milou, Taco and Victor. They listened to my moaning on a 24/7 basis and provided me mental support, jokes and candy. Thanks especially for always putting things into perspective and laughing at me and your absurd selves.

Finally, I want to thank professor A. White, specialized in academic English en natuurlijk papa en mama. Zou meneer Ossewaarde het geloven als hij me zag?

Contents

1. Introduction

The semantics of modified numerals is a widely debated topic with many unanswered questions (Geurts and Nouwen 2007; Büring 2007; Cummins and Kastos 2010). One of these questions concerns the unavailability of Scalar Implicatures (SI's) for constructions containing the comparative modifier *more than.*

(1) Barbara read 9 books.

SI: ¬Barbara read 10 books

Predicted interpretation: ✔ Barbara read exactly 9 books.

(2) Barbara read more than 9 books.

SI: ¬Barbara read more than 10 books

Predicted interpretation: **X** Barbara read exactly 10 books.

It's remarkable that the inference that is derived from (1) is not available for (2). The inference in (1) relies on the maxim of quantity (Grice, 1991). In the assumption that the speaker is as informative as possible, a hearer will reason that if Barbara read more than 9 books, the speaker would have said so, and thus concludes that Barbara read exactly 9 books. In other words, by using the weaker number, the speaker denies the stronger number (Horn 1984; Levinson 2000).

There is no a priori reason to expect that (2) would not trigger a similar implicature, and therefore the interpretation of (2) is surprising. In addition, the implicature can be generated when the modified numeral is embedded under a universal quantifier (Fox and Hackl 2006; Mayr, 2015).

(3) Every student read more than 9 books.

SI: ¬Every student read more than 10 books.

SI infers: Some student read exactly 10 books.

Predicted interpretation: ✔ Some student read exactly 10 books.

The lack of SI in (2) and its availability in (3) have left the field of modified numerals with a complex puzzle. The aim of this thesis is to enrich this research field with quantitative data. The structure of the thesis will be as follows. First, I will discuss the theoretical background of the observations in (2) and (3).

Second, I will show that processing experiments have provided insight in theoretical semantic issues. These two chapters lead to an experiment, that I will describe in chapter 4. Finally, I will discuss my findings in chapter 5.

2. Modified numerals

For (1), a sentence containing a bare numeral, I showed that the inference that Barbara did not read more than 9 books is attributed to the speaker, which leads to an exact interpretation of the numeral. In contrast, we saw that there is no such an inference on (2), an utterance containing the comparative modifier *more than*. The difference between (1) and (2) is unexpected, in that way that the mechanism that explains (1) – the maxim of quantity - is expected to be applicable to (2). There are two main theories that aim to account for this difference. In this chapter, I will discuss these theories and the questions they raise.

2.1 Density

The first theory I will discuss assumes that all scales of degree are dense. In this section, I will discuss how this explains the observations in (1), (2) and (3), what support there is to believe that scales are dense, and what the problems are with the theory of density.

2.1.1 The UDM

Fox and Hackl (2006) propose that the scales of all degree domains are always dense.

(4) The Universal Density of Measurements (UDM): Measurement scales that are needed for natural language semantics are *always* dense. (Fox and Hackl 2006, p. 542)

This means that that for any two degrees, *d* and *e*, there exist infinitely many degrees between *d* and *e*. This predicts the lack of SI's in (2). Let us assume the following semantics.

(5a) Barbara read more than 9 books.

(5b) max*d*(Barbara read *d*-many books) > 9 (Nouwen 2008, p. 281)

There are two underlying assumptions for (5b), the first assumption is that the *more than* construction has a maximality operator, an operator that takes degree, or cardinality properties and returns their maximal member. A second assumption is that it contains a silent *many* (Hackl 2000). Hence, the comparative modifier applies to a set of degrees *d* such that there is a plurality of books *x* that Barbara read and the cardinality x ≥ *d*.

For a potential implicature, the stronger alternatives will be negated. Thus, the first things that need to be calculated are these stronger alternatives. Arguing from a standard account, these would be of the form in (6).

(6) max*d*(Barbara read *d*-many books) > *n* (Nouwen 2008, p. 281)

In a traditional account of SI's, *n* would be any natural number exceeding 9. In that case, a SI can be calculated in the same way as for bare numerals. However, for a dense scale *n* will be anything between 9 and ∞. This leads to the potential implicatures in (7).

(7) ¬[max*d*(Barbara read *d*-many books) > *n*] and *n* is in (9,∞) (Nouwen 2008, p. 281)

A problem occurs when deriving these implicatures. The assertion (5b) requires a *d* such that Barbara read 9+*d* books. Because the scale is dense, for every *d* there is an implicature that denies Barbara read 9+*d*/2 books. If the implicature mechanism wants to negate all the stronger alternatives, it contradicts the assertion on (5b). Thus, it is not possible to calculate an implicature without contradicting the assertion. Because of this contradiction implicatures will not be calculated. This explains why (2) is interpreted logically. The contrast between the assertion and the totality of the implicatures does not allow deriving the implicature. (Fox and Hackl 2006)

2.1.2 Universal Modals

In (3), I showed that a SI can be generated when the modified numeral is embedded under a universal quantifier. This is the same for universal modals, but not for existential modals. This observation is predicted by the UDM. To explain this, Fox and Hackl (2006) make use of the operator *only*. It is generally known that the SI of a sentence can be made explicit by the use of *only* (Groenendijk and Stokhof 1984).

(8) The only implicature generalization (OIG): Utterance of a sentence, S, as a default, the inference/implicature that (the speaker believes) onlyS', where S' is (a minimal modification of) S with focus on scalar items. (Fox and Hackl 2006, p. 6)

This generalization is illustrated in (9). In (9a) the SI that is drawn from the use of *some*, is made expressed literally. In (9b) this implicature is already expressed using the operator *only*, and therefore cannot be expressed literally again.

(9a) John found some of the marbles; in fact he found all of them.

(9b) John only found some of the marbles; #in fact, he found all of them.

Keeping this in mind, let us take a look at the following two examples in which the modifier *more than* is embedded under a modal verb.

(10) Barbara is required to read more than 9 books

SI: ¬There is a degree d greater than 9, such that you are required to read more than d books

Predicted interpretation: ✔Barbara is not required to read more than 10 books

(11) Barbara is allowed to read more than 9 books

SI: ¬There is a degree d greater than 9, such that you are allowed to read more than d books

Predicted interpretation: **X** Barbara is not allowed to read more than 10 books

The OIG predicts that the interpretation of (10) corresponds to the statements containing *only* in (10'). This prediction is correct. Further, the prediction is that the interpretation of (11) corresponds to the meaning of (11'). However, (11) cannot be interpreted with a SI. This is illustrated in (11'). The interpretation in (11) cannot be expressed using *only*, because there is no SI that only can express. This leads to the infelicity in (11').

(10') Barbara is only required to read more than 9 books.

(11') *Barbara is only allowed to read more than 9 books.

Generally, these examples show that the universal modal *required* allows for generating a SI, whereas the existential modal *allowed* does not.

The difference between (10) and (11) is predicted by the UDM, and is related to the semantic properties of the modal operators *required* and *allowed*. Let us first consider *allowed*. Assuming the semantics described above, the assertion for (11) is that there is a world *w* in which Barbara read 9+*d* books. The potential SI is that there is no world such that Barbara read more than 9+*d*/2 books. Assuming a dense scale, there is an implicature contradicting the assertion for each possible *d*. The contradiction between the assertion and the implicatures leads to a logical interpretation of the sentence.

For *required*, such a contradiction does not occur. The assertion is that Barbara is required to read 9+*d* books. For each possible *d*, there is a world *w* in which this is the case. There is no contrast with the implicature in (10) and the SI can be derived. More detailed, for every world *w*, there is degree *d* such that Barbara read 9+*d* books in *w*. Reasoning from the UDM, this also means that there is a degree 9+*d*/2 between 9 and 9+*d*, such that Barbara read more than that degree of books in *w*. However, an existential and a universal quantifier cannot commute without effecting truth conditions. Therefore, it cannot be concluded that there is a degree *d* greater than 9 such that Barbara is required to read more than *d* books. Rather than a dense set of degrees, the modal operates over a dense set of accessible worlds corresponding to the degrees. In this way, there is no contradiction, and the SI can be derived.

In chapter 4, I will present an experiment that investigates the interaction of modified numerals and universal quantifiers, but instead of modal quantifiers, I will use nominal ones. It is not entirely clear whether Fox and Hackl's (2006) and Mayr's (2015) reasoning about modals is fully transferable to the nominal domain, although the logical structure would certainly appear to be the same.

2.1.3 Supporting the UDM: no more than

In the previous section, I showed that the UDM predicts the lack of SI's for modified numerals and their availability under universal modals and universal quantifiers. In addition, Nouwen (2008) shows that the UDM also correctly predicts the presence of SI's under non-strict comparison. Let us consider (12) and its semantics in (13).

(12) Barbara read no more than 9 books.

(13) max*d*(Barbara read *d*-many books) $\leq$ 9 (Nouwen 2008, p. 276)

From (12) it follows that the total of books Barbara read is smaller than or equal to 9. The preferred interpretation however, is that the total books equals 9. To explain this, let us consider the stronger alternatives of (12) in (14).

(14a) max*d*(Barbara read *d*-many books) $\leq$ 8

(14b) max*d*(Barbara read *d*-many books) $\leq$ 7

(14c) max*d*(Barbara read *d*-many books) $\leq$ 6

(14d) max*d*(Barbara read *d*-many books) $\leq$ 5

…………………. (Nouwen 2008, p.278)

The conjunction of the assertion in (12) and the negation of its stronger alternatives in (14) lead to the interpretation in (15).

(15) max*d*(Barbara read *d*-many books) = 9

The potential implicature for (12) can be formulated as follows.

(16) $\neg[\max d(\text{Barbara read } d\text{-many books}) \leq n]$ and $n \in (9,\infty)$

This means that there is a degree $d \in (9,\infty)$ such that Barbara read 9+*d* books, and for every $d > 0$, there is an implicature that denies that he read 9 + *d*/2 books. In this case however, contrary to what we saw for unembedded modified numerals, there is no implicature denying the assertion that Barbara read 9+0 books. If the implicature is accepted, the interpretation is that Barbara read exactly 9 books, which does not contradict the assertion. Because there is no contradiction, (12) can be interpreted exhaustively.

2.1.4 Problems with the UDM

In the previous section, I showed that the UDM predicts the lack of implicature for the comparative modified numeral *more than*. In addition, the UDM also predicts the lack of implicature for *more than* when embedded under existential modals and existential quantifiers. Finally, it accounts for the fact that SI's are available under universal modals, universal quantifiers and under non strict comparison. However, there are two problems with the UDM. First, it does not make the right predictions for the superlative modifier *at least*. And second, it does not correctly predict the availability of SI's under negation. In the following sections, I will discuss these problems.

2.1.4-a At least

The UDM does not predict the lack of SI's for the superlative modifier *at least*. From a naive perspective, *at least* seems to be corresponding to *no more than*, in the sense that is expresses non-strict comparison (≥), and thus the availability of an SI is predicted. In the previous section, I showed that non-strict comparison allows for the derivation of a SI. Based on this, the expectation is that this is the same for *at least*. This is contradicting the data, as is shown in (16).

(16) Barbara read at least 10 books

SI: ¬Barbara read more than 10 books.

Predicted interpretation: **X** Barbara read exactly 10 books.

It needs to be noted that Fox and Hackl (2006) are aware of this fact and do not aim to account for the availability of SI's for *at least*. They claim that an alternative account for *at least* is required. Nouwen (2008) provides such an account by stating that we cannot analyze *at least* as a modifier that expresses non-strict comparison. There is more going on in its semantics than the modification of a numeral. *At least* expresses the speaker's ignorance about the exact number. Consider the following example.

(18) Mark had at least 5 pints.

From (18), a hearer normally concludes that the speaker does not know the exact number of beers Mark consumed. Geurts and Nouwen (2007) argue this is due to the specific semantics of *at least*. In its semantics, *at least* expresses modal attitudes to alternatives on scales. Thus, rather than only non-strict comparison (≥), *at least* corresponds to the following (Geurts and Nouwen 2007).

(19) Mark had at least 5 pints.

□[Mark had five pints] & ◊[Mark had more than five pints]

By default, the modality expressed by *at least* is epistemic (Geurts and Nouwen 2007). In (18) it is explicitly stated that the speaker considers the possibility that Mark had more than 5 beers. If the speaker would know exactly how many pints Mark has consumed, (18) would be infelicitous. If it is likely that the speaker does not know the exact number of pints, this would render an utterance of (18) acceptable.

Nouwen (2008) explains the lack of implicature by the epistemic modality of *at least*. Its semantics are comparable to that of *only* or the exhaustivity operator. In their semantics, superlative modifiers use up the alternatives (Geurts and Nouwen 2007). This means that the lack of implicature for *at least* is not problematic for the UDM, as long as an alternative account is considered.

What is problematic for such an account, is that *at least* seems to allow for SI's in certain linguistic contexts. Assuming at leasts epistemic modality, no SI's are expected in any linguistic context. However, embedding under a universal modal allows the derivation of SI's, as is shown in (20).

(20) Barbara is required to read at least 9 books

SI: ¬There is a degree *d* greater than or equal to 9, such that you are required to read d books

Predicted interpretation: ✔Barbara is only required to read at least 9 books

Just as for *more than*, embedding under a universal modal licenses the possibility to derive a SI. This is unexpected from the perspective that at least uses up its alternatives, because of the modality in its semantics. That perspective predicts that there is no linguistic context in which SI's are available for *at least*.

2.1.4-b Sentential negation

Above, we saw that the negation of *more than* leads to non-strict comparison, and hence to the availability of an SI. However, Mayr (2015) pointed out that that *more than* lacks an SI under sentential negation. Let us consider (21).

(21) Barbara didn't read more than 9 books

SI: ¬ For all degrees *d* smaller than 9 it is the case that Barbara did not read more than *d-many* books

SI: For all degrees *d*, smaller than 9, it is the case that Barbara read *d-many* books

Predicted interpretation: ✗ Barbara read more than 8 books

The assertion of (20) can be formulated as follows.

(22) max*d*(Barbara read *d*-many books) ≤ 9

The assertion is that Barbara read 9 books or fewer. The potential implicature is that for all degrees smaller than 9 it is the case that Barbara read that amount of books. It is obvious that the assertion and the implicature are not contradicting. However, the SI is not generated, and the sentence is interpreted logically (Mayr,205). This is problematic for the UDM, and this needs to be explained. The second theory that I will discuss, does aim to account for this difference.

2.2 Non-monotonic Alternatives

To account for the observations in (1), (2) and (3), the second theory assumes lexical alternatives for constructions like *at least n* and *more than n* and is proposed by Mayr (2015). The semantic differences between *at least* and *more than* as analyzed by Geurts and Nouwen (2007) are not relevant for this theory.

2.2.1 Alternatives

The theory relies on the assumption that for both modifiers there are lexical alternatives. These alternatives are not ordered by monotonicity, which means there are no entailment relations between the alternatives. The relevant Horn sets are formulated as in (23). These sets lead to the alternatives as in (24) and (25).

(23a)Alt(at least)=Alt(at most) = {at most, at least}

(23b) Alt(more than)=Alt(fewer than) = {fewer than, more than}

(23c) Alt(n) = {…n-1,n,n+1…}

(24a) More than 3 boys left.

(24b) Alt([[More than three boys left]]) = {more than 4 boys left, more than 5 boys left, . . ., fewer than 4 boys left, fewer than 5 boys left, . . .} (Mayr 2015)

(25a) At least 4 boys left.

(25b) Alt([[At least three boys left]]) = {at least 3 boys left, at least 4 boys left, . . ., at most 3 boys left, at most 4 boys left, . . .} (Mayr 2015)

Mayr (2015) shows that in the cases where no SI can be derived, there is a symmetry problem by employing the alternatives. He assumes an exhaustivity operator that only negates alternatives whose negation does not automatically require the truth of some other alternative. He refers to those alternatives as the *innocently excludable* ones, following Fox (2007). The negation of those alternatives is not part of the strengthened interpretation. Such a mechanism requires a search for the innocently excludable alternatives. In (25b), there are no innocently excludable alternatives. In addition, the negations of the alternatives in (25b) contradict each other. This leads to the lack of a SI.

The same is applicable to the alternatives for *more than n* in (24a). Although *fewer than 4 boys* can be negated, this is in line with the logic interpretation. Negation of this alternative forces inclusion of the assertion, which is innocent. Thus, in the same way as for *at least*, no SI is generated for (23a) and the sentence will be interpreted logically.

This analysis also explains the availability of SI's under universal nominal quantifiers. Consider (26).

(26a) Every student read at least 9 books.
(26b) Alt([[Every student read at least three books]]) ={Every student read at least 9 books, Every student read at least 10books, . . ., Every student read at most 9 books, Every student read at most 10 books, . . .}

The alternatives that contain *at most* are innocently excludable. The negation of the alternative *Every student read at most 9 books* is not part of the strengthened interpretation. In contrast to what we saw above, the negation of *at most 9* does not lead to inclusion of *at least* 4. If some students read more than 9 books, it does not follow that all students read more than 9 books. This would be required by inclusion of the alternative *every student reads at least 10 books.* This means the implicature can be derived, which corresponds to the actual interpretation of the sentence. This model also correctly predicts the availability of SI's *more than* under a modal verb for (Mayr 2015).

2.2.2 Supporting non-monotonic alternatives: Mayr's generalization

Support for the account of monotonic alternatives comes from the contexts in which SI's are available for modified numerals. According to Mayr (2015) the availability of SI's for modified numerals is depending on specific linguistic contexts. These contexts are the same for *at least* and *more than*.

Table 1. Mayr 2015: Contexts that license SI's for comparative and superlative modified numerals

	root	¬	∀	□	∃	◊	¬∃	>1/2	∧	∨
More than n	-	-	+	+	-	-	+	+	+	-
At least n	-	-	+	+	-	-	+	+	+	-

There is a systematic difference between the context that license SI's and the contexts that do not license SI's. In the contexts that do not allow SI's (¬, ∃, ◊, ∨) the strengthened interpretation of the sentence – i.e. the interpretation after generating a SI - would have been entailed by the logical interpretation of a minimally differing sentence with an *exactly n* expression. Let us consider (27) and (28). The arrow ($\rightsquigarrow$) expresses non-entailment inferences.

(27a) Every student read more than 9 books. $\rightsquigarrow$ Not every student read more than 10 books.

(27b) Every student read exactly 10 books. $\not\rightsquigarrow$ Not every student read more than 10 books.

(28a) Someone read more than 9 books. $\not\rightsquigarrow$ No student read more than 10 books.

(28b) Every student read exactly 10 books. $\rightsquigarrow$ No student read more than 10 books.

If the modified numeral in (27) is replaced by an *exactly n* expression, its logical interpretation would be the proposition that every student wrote exactly 10 books. This does not infer the strengthened interpretation (27). In (28), the strengthened interpretation *no student read more than 10 books* follows from the logical interpretation of the sentence with *exactly*. Based on this observation, Mayr (2015) makes the following generalization.

Mayr's Generalization (Mayr 2015, p. 150)

For any φ with a modified numeral α n in it, a scalar implicature for α n is generated iff [[ψ]] does not entail [[φ]] S, where ψ is just like φ except that α n is replaced by exactly n+1, exactly n−1,or exactly n.

This generalization corresponds to the predictions of the theory described above. This means that the second theory makes the right predictions for both *more than* as *at least* in various different linguistic contexts. The question is however, whether the generalization can be supported by quantitative data.

2.3 Questions

In the previous sections I discussed two theories on the availability of SI's for modified numerals. The UDM assumes density for all degree scales and successfully predicts the lack of implicature for *more than* and the availability of SI's for *more than* embedded under universal modals. However, it does not predict the lack of implicature for *more than* under sentential negation. In addition, it requires an alternative account to explain the lack of implicature for *at least*. Such an account is proposed, however, this theory does not account for the availability of SI's for *at least* in the environments in table 1.

The second theory assumes lexical, non-monotonic alternatives for *at least n* and *more than n*. This account successfully predicts the availability of SI's for both modified numerals in certain linguistic environments. Further, it does not predict a difference between *at least* and *at all*, in terms of the availability of SI's.[1] .

At this point, there are three main questions. The first two questions concern the availability of SI's in specific linguistic contexts, and the third one is related to the difference between *at least* and *more than*. In this section, I will motivate these questions and show how an experiment can contribute to answering them.

2.3.1 Scalar Implicatures

The first question is whether it is true that SI's can be derived from modified numerals when embedded under universal nominal quantifiers[2]. Let us reconsider (3).

(3) Every student read more than 9 books.

SI: ¬Every student read more than 10 books.

SI infers: Some student read exactly 10 books.

Predicted interpretation: ✔ Some student read exactly 10 books.

---

[1] Please note that the theory and the generalization are two different things. The theory predicts the generalization. However, the generalization can be explained in other ways. See Mayr (2015) for another theory that possibly explains the generalization

[2] However, further investigation of the other linguistic contexts in table 1 is recommended for further research.

In the introduction, I claimed that the predicted interpretation corresponds with the actual interpretation. This claim is shared by Mayr (2015) and to a lesser extent by Fox and Hackl (2006). They expressed their uncertainty about this intuition (Fox and Hackl 2006, pp. 52). It is important to know whether it is true that SI's for *more than* are available under universal nominal quantifiers. The reliability of a theory will increase if quantitative data shows this is indeed the case.

Obviously, it needs to be investigated whether this is testable. The first step in order to do so is to see whether speakers derive the inference that is predicted. Such an investigation can be realized by testing an underinformative sentence. Bott and Noveck (2006) for example, conducted experiments using underinformative sentence with the scalar item *some*. If a hearer hears a sentence containing the item *some* that corresponds to a situation in which the stronger alternative *all* could have been used, this sentence can be judged infelicitous. A truth-value judgment task could be the first step in investigating the pragmatic inferences of sentences like (3).

This leads us to the second question. If there is a pragmatic inference, is it true that it is a SI? In order to answer this question, processing data is extremely useful. It is known, as I will elaborate on in the next chapter, that SI's are costly to process. Previous experiments have shown that the rejection of an underinformative sentence – which means an SI is derived- is associated with longer reaction times than the acceptance of such a sentence. For modified numerals, this means that longer reaction times for responses that rely on an inference, indicate that these inferences are SI's.

2.3.2 Two types of modifiers

The third question concerns the availability of SI's for both *more than* as *at least*. There are two reasons for asking this question. First, it is known there are semantic differences between *more* and *at least* (Geurts and Nouwen 2007). Above, I discussed for example that at least expresses speaker's ignorance. It is possible that these different semantics lead to a difference in the availability of the SI's as well. This brings us to the second reason. The UDM relies on an alternative account to explain the lack of implicature for *at least*, whereas the non-monotonic alternatives account makes no distinction between the two modifiers. If there is a difference, this would contradict Mayr's (2015) statement that at least and *more than* can be analyzed in the same way. If there is no difference between the two modifiers, this supports Mayr's statement that the availability of SI's for *at least* and *more than* can be analyzed in the same way.

Again, such a difference is testable in a processing experiment. First, two types underinformative sentences, as explained above, can be designed. If one contains a superlative modifier and the other contains a comparative modifier, it allows us to compare them in terms of responses. Second, processing data can be valuable to trace a potential difference. Previous studies have showed differences between superlative and comparative modifiers in processing terms. This means that if there is a difference, this will possibly show in a reaction time experiment.

Thus, what is required in order to answer these questions is first, responses to underinformative sentences containing a modified numeral (*at least* or *more than*), and second, processing data of the responses to these sentences. In the following chapter, I will elaborate on processing SI's and modified numerals.

3. Processing Modified Numerals and Scalar Implicatures

Processing data can help us answering the research questions possessed above. In this chapter, I will discuss how processing data has contributed to questions in the field of modified numerals and scalar implicatures, and what this means for the questions of this thesis. This chapter consists of three sections. First, I will discuss what is known about the processing profile of modified numerals. More specifically, I will show in what way processing data has contributed to semantic puzzles. Second, I will do the same thing for SI's. The interest of this thesis is in modified numerals embedded under a universal nominal quantifier. For this linguistic context, the availability of SI's is predicted. As I stated above, SI's have a specific processing profile. In the second section, I will elaborate on this. Based on the first two sections, and the theoretical chapter above, I will formulate hypotheses for the questions above.

3.1 Processing modified numerals

The third question of chapter 2 concerned the possible difference between *more than* and *at least*. Differences between the two modifiers can be observed in terms of their cognitive profile. This difference can be observed on two levels. First, it has been observed that there is delay in acquisition for *at least*, and second, *at least* displays a delay in online performance. These two observations suggest that *at least* has a complexity in its semantics that is not associated with *more than*. In this section, I will explain these observations and their implications for the questions of this thesis.

3.1.1 Delay in Acquisition

It is found that *at least* is required at a later age than *more than*. Musolino (2004) compared the cognitive profile of superlative modifiers to that of comparative modifiers. He showed that children have difficulties with tasks that include superlative modifiers, whereas they successfully complete tasks that involve comparative modifiers. He conducted an experiment in which participants got to see cards representing zero to four objects. They were asked to pick that card that represented *Q n* objects, where *Q* was either *exactly n*, *more than n*, or *at least n*. Musolino (2004) compared the results of children (age 4-5) to that of adults. Adults performed close to perfectly in all conditions. The children group however, was 100% accurate on *exactly n*, 88% accurate on *more than* and at chance on *at least*. These results demonstrate that the children have no full understanding of the meaning of superlative modifiers.

Geurts et al. (2010) carried out a comparable experiment. They presented participants with sets of six boxes, some of which were filled with an object. The participant's task was to use the boxes and the objects to represent the sentence they listened to. They could either add objects, remove them or leave them as they were. The sentences were of the form *Q n of the boxes have a toy*, where *Q* was the same as in Musolino's (2004) experiment. Children performed 97% on *more than*, and 88% on *at least*. What was especially remarkable is that children performed only 42% correctly on *at most*. Again, children had more difficulty performing tasks involving superlative modifiers than comparative modifiers. The fact that children performed 88% in the experiment of Geurts et al. (2010) and at chance in the experiment of Musolino (2004) may be due to the different age groups participating in their experiments. Musolino (2004) conducted the experiment on 4-5 year olds, whereas Geurts et al. (2010) tested 11 year olds.

The results of both experiments show that the meaning of comparative modifiers is understood earlier in life than the meaning of superlative modifiers.[3] This suggests a complexity in the meaning of *at least* that cannot be associated with *more than*.

3.1.2 Reaction Times

In addition to the acquisition study described above, Geurts et al. (2010) conducted an offline task, consisting of a questionnaire that participants were asked to fill in. The items were single-premiss arguments, participants had to indicate whether the conclusion followed from the premiss, by ticking a box. The premiss-conclusion combination of the critical items were either *X had n drinks* → *X had at least n drinks*, *X had n drinks* → *X had at most n drinks*, and *X had at most n drinks* → *X had at most n+1 drinks*. The control items were of the same format, but with a comparative modifier instead of a superlative. Geurts et al. (2010) found that constructions with comparative modifiers were endorsed more often than their superlative counterparts. However, participants produced correct responses on all trials. This is in different from the data we have for children and suggests that, once their correct meaning is acquired, all modifiers are equally difficult to understand.

[3] Obviously, this fact is extremely valuable for analyses of the semantic nature of *at least*. The delay in acquisition suggests a more complex nature of at least. This complexity can be explained in different ways. On the one hand, it can be related to the account that superlative quantifiers are semantically modal (Geurts and Nouwen, 2007). This is the same account I discussed in the introduction and that predicts no SI's on superlative quantifiers. On the other hand, the pragmatic account of superlative quantifier meaning (Cummins and Katos, 2010) predicts a similar delay for different reasons. For this section, I will leave this debate and simply accept the complex nature of *at least*, compared to *more than*.

To test this, Geurts et al. (2010) carried out an experiment that focused on the processing profile of the two types of modifiers. First, they presented a statement of the form *There are at least n A's*. Next, they unveiled a situation consisting of certain letters. The task of the participant was to judge whether these statements were correct in combination with the situation. To investigate the processing properties, both the reading time and the reaction time were measured. The trials were separated into a reading stage and a reaction stage, because of the possible relation between these stages and distinct cognitive processes. First, a sentence needs to be interpreted (reading time) and second its interpretation must be confronted with the facts (Geurts et al., 2010). There were no differences in the reading time, but there was an effect on the reaction time[4]. Participants responded significantly slower to superlative modifiers than to comparative modifiers.

Finally, Cummins and Katsos (2010) found a similar effect for non-linguistic expressions of comparison. They conducted an experiment similar to Geurts et al. (2010). Instead of a sentence, they showed participants a formula of the form *X ? N*, where X denotes a letter (A or B), N a numeral and ? a symbol (either =, >,<, ≥ or ≤). The outcome was similar to the outcome of Geurts et al. (2010). More specific, the response times for ≥ (corresponding to *at least*[5]) were longer than the response times for > (corresponding to *more than*[6]). This shows that, apart from the durability of different accounts on the complexity of *at least*, we know that in itself, non-strict comparison is costly. It takes more time to process an inclusive construction (*at least*) than an exclusive construction (*more than*).

Please note that at this point, the semantic nature of *at least* is not an urgent issue for this thesis. Both the view that superlative modifiers are semantically modal (Geurts and Nouwen, 2007), and other accounts on the semantics of superlative modifiers (Cummins and Katos, 2010) predict a delay in acquisition and a delay in processing time. The relevant point for this thesis is that it is known that *at least* and takes longer to process than *more than*.

Recall that the question of this thesis is whether there is a difference between *at least* and *more than* in their ability to generate an SI when embedded under a universal nominal modal. Above, I stated that I will investigate this by responses and that processing data could contribute to this question. In this section, I

[4] In addition, Geurts et al. found an effect of entailment direction, which was attested in the literature before (Geurts and van der Slik, 2005). In this thesis however, I will focus on *at least* and *more than*. The monotonicity properties of these two modifiers are the same and therefore, the effects of monotonicity properties on processing are not relevant for this thesis. This does not take away the importance and relevance of the processing effects found for entailment direction.

[5] ≥ only corresponds to *at least* from a naïve perspective.

[6] > only corresponds to *more than* from a naïve perspective.

showed that differences in reaction times can be observed for the two modifiers. This means that differences between *more than* and *at least* can possibly be traced by measuring reaction times.

3.2 Processing Scalar Implicatures

In the previous section, I showed that a difference between *more than* and *at least* can be found in processing data. In order to answer the research questions the possibility of tracing SI's in an experimental setting needs to be explored as well. In this section, I will discuss the question whether it is possible to identify SI's by analyzing processing data.

3.2.1 Pragmatic responses are slower than semantic responses

SI's have a clear processing profile as is shown by Bott and Noveck (2004), who conducted an experiment on the SI that is derived from expressions that contain the quantifier *some*. Their experiment consisted of a truth value judgment task for different types of sentences containing a scalar term. The critical items contained *some* and were underinformative in that way that they were semantically correct, but that a stronger term could have been used. An example is *Some elephants are mammals*, from which the SI *but not all* can be drawn. Participants were asked to judge the sentences *true* or *false*. For these critical items, a *false* response is associated with the interpretation *some but not all*, the interpretation after a SI is generated. I will refer to these responses as a pragmatic. A true response is associated with the semantic interpretation of *some*, including the possibility of *all*. I will refer to these responses as semantic.

The majority of the participants responded mixed to the critical items, which means they responded sometimes pragmatically and sometimes semantically. Bott and Noveck (2004) compared the semantic responses to the pragmatic responses and found that the RT's for pragmatic responses were longer than the RT's for semantic responses. In addition, the RT's for pragmatic responses were slower than the RT's for all of the control sentences. To control for a possible delay for rejection compared to approval, the control items included items that required false responses. The pragmatic responses for the critical items were significantly slower than responses for these control items. This shows that a pragmatic response to underinformative sentences is characteristically different from all the other items. The RT's for logical responses were similar to the RT's for the control items. Bott and Noveck (2004) therefore consider a logical response to be the immediate response to underinformative items. To derive a SI, an extra step needs to be made on top of this immediate response. This step appears to be costly. In addition to this, Bott and Noveck (2004) conducted a second experiment, in which participants performed the exact same task under time pressure. They found that when participants were given limited time to response, the amount of logical responses increased. This strengthens the idea that processing a SI is costly. If participants do not have enough time to derive an SI, they are less likely to do so.

For the questions of this thesis, this means that it is possible to identify SI's in an experimental setting. If a hearer derives an inference for an underinformative sentence and the RT is longer than for a semantic response, this suggests the inference is a SI. Thus, if people draw SI's from modified numerals under universal quantifiers, the expectation is that these responses have longer RT's than semantic responses.

3.3 Hypotheses

Let us return to the questions possesed in chapter 3. The first two concerned the availability of SI's for modified numerals embedded under a universal quantifier. The third question concerned a potential difference in the availability of SI's for *more than* and *at least*. I stated that processing data could contribute to answering these questions. In this chapter, I elaborated on this and showed that RT data can possibly be valuable in order to answer this question. In this section, I will discuss how this leads to an experiment and provide an hypothesis.

3.3.1 Scalar Implicatures

The question of this thesis is whether SI's can be derived from modified numerals when they are embedded under universal nominal quantifiers. In this chapter, I showed that deriving a SI is associated with a longer RT, as showed by Bott and Noveck (2004). In order to investigate in SI's for modified numerals, their design can be used. In the same way as they did, a truth-value judgment task can be conducted. The task has to consist of a sentences that are underinformative for the context they are associated with. Consider (29)

(29) Every bowl contains more than 3 marbles

SI: ¬Every bowl contains more than 4 marbles

Predicted interpretation: + There is at least 1 bowl that contains exactly 4 marbles

If a SI is derived, (29) is not compatible with a situation where all bowls contain 4 marbles or more. In an experimental setting, such a context can be created by demonstrating a picture combined with a sentence. Just as in the design of Bott and Noveck (2004) participants will have to be asked to judge the acceptability of the sentence. Rejection of the sentence means that an inference is derived, whereas acceptance means that the sentence is interpreted logically. To see whether the processing profile of the inference that the rejection is based on corresponds to that of the SI's, RT's will be measured.

Both the UDM (Fox and Hackl 2006) and by the theory of non-monotonic alternatives (Mayr 2015) predict the availability of SI's for *more than* when they are embedded under universal quantifiers. For the experiment, this means the expectation is that the responses are mixed and that pragmatic responses are slower than semantic responses.

3.3.2 Modifiers

The second question is whether there is a difference in the availability of SI's between *at least* and *more* than. For the methodology, this means that two types of sentences need to be considered, sentences containing *more than* and sentences containing *at* least. This allows us to compare the two groups in terms of responses and RT's.

For the modifiers, the expectations are less obvious. The non-monotonic alternatives theory predicts no difference between *at least* and *more than*. For both modifiers, the expectation is that participants give mixed responses and that the pragmatic responses are slower. From the UDM and theories on *at least* the expectations are not entirely clear. Fox and Hackl (2006) expressed their uncertainty about the availability of a SI under a universal nominal quantifier, but theoretically the prediction is that it corresponds to the availability of SI's for universal modals. Moreover, The UDM only predicts a SI for *more than* and makes no statements about at least. However, a difference between *at least and more than* would not be unexpected from that perspective. The additional theory on the semantics of at least (Geurts and Nouwen 2007; Nouwen 2008) consider *at least* to be a different creature than *more than* by nature. Considering the fact that it is known that the semantic properties of the two modifiers differ in many aspects, it would not be surprising if they differ in this aspect as well.

Arguing from both perspectives it is expected that the RT's for *at least* are in general longer than the RT's for *more than*. This is predicted by Geurts et al. (2010) who showed *at least* has a heavier processing load than *more than* in general. It is likely that a similar effect can be observed for the sentences of the experiment.

4. The Experiment

In this thesis, we want to investigate (a) whether it is possible to obtain quantitative data that shows that scalar implicatures (SI's) on modified numerals can be generated when they are embedded under universal nominal quantifiers. In other words, we want to investigate whether the inference in (29) can be traced in an experimental setting.

(29) Every bowl contains at least 3 marbles. □ Not every bowl contains at least 4 marbles.

Further, we want to investigate (b) whether superlative and comparative modifiers show differences within this linguistic context, and (c) whether it is the case that the SI is generated, whether it displays the same processing profile as well-known SI's - like *some* leading to the implicature *but not all*– in the case of generating. As I discussed in chapter 3, Bott and Noveck (2004) showed that deriving a scalar inference is costly, and that participants respond slower when reasoning pragmatically than when reasoning semantically. For the research topic of this thesis, a similar effect is expected. To research this, we designed an experiment similar to Bott and Noveck's (2004). The task is as described above. Participants are asked to judge whether they found a sentence picture combination natural or unnatural. The target sentences consisted of expressions including a modified numeral embedded under a universal nominal quantifier. For the critical items, the sentences were a underinformative description of the picture. For example, for the sentence in (29) the picture showed three bowls, each containing four marbles or more. A SI response leads to a negative response (*unnatural)* whereas a semantic interpretation leads to a positive response (*natural*).

4.1 Methods

The experiment employed a truth value judgment task in which participants were asked to judge sentence-image combinations. The combinations of the target items were semantically a good match, but allowed, theoretically, for a SI that would lead to an *unnatural* response. If the SI is generated, participants will judge the combination to be *unnatural*. SI's are not always mandatory. Therefore, we expect mixed responses: sometimes *natural* and sometimes *unnatural*. If the responses are indeed mixed, we expect the reaction time for *unnatural* responses to be longer. This prediction is in line with results from previous studies (Bott and Noveck, 2004) in which semantic responses were always quicker than pragmatic responses.

4.1.1 Materials

The materials consisted of 126 items, of which 42 were test items, and 84 were fillers. Both target and filler sentences were combined with an image of 3 bowls filled with a certain amount of marbles. The test sentences always contained a numeral modified by either a comparative or a superlative modifier. The numeral in the sentences varied from 3 to 9. The experiment was in Dutch, but both in the description of the methodology and in the discussion of the results, we will often use English translations to refer to the materials. We distinguished three types of test items, as illustrated in table 1. There are three types of test items: pragmatic, semantic and false. The type of test items is determined by the image that it is combined with. For the first type (pragmatic), the test sentence is combined with a picture that semantically matches the picture, but pragmatically does not. With the latter we mean that the sentence-image combination

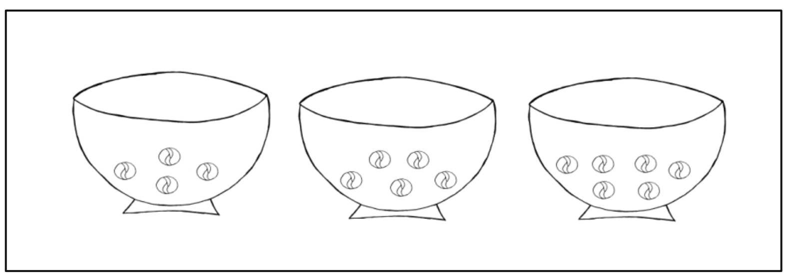


Figure 1 Target item for the pragmatic condition, when showed in combination with 'Every bowl contains at least 3 marbles'.

allows for the derivation of a SI is derived. In addition, for the second type (semantic) the combination is both semantically and pragmatically true. For the third type (false) the combination between the sentence and the image is incompatible from both a semantic and a pragmatic perspective. To describe the picture in the tables –and in this thesis, I use the notation (n)(n)(n), in which every () represents a bowl, and the n represents the number of marbles in the bowl.

Table 2 Overview target items

Reference	Example Sentence	Example Picture Description	Expected Response
pragmatic	Every bowl contains at least 3 marbles	(4) (5) (6)	natural/unnatural
semantic	Every bowl contains at least 3 marbles	(3) (4) (5)	natural
false	Every bowl contains at least 3 marbles	(2) (3) (2)	unnatural

Every participant was exposed to fourteen trials of each type, of which seven were combined with a comparative modifier and a numeral in the range 3-9, and seven contained a superlative modifier and a numeral in the range 2-8.

The filler items consisted of similar images –bowls with various amounts of marbles- and three types of sentences. In the first type (general), the images were combined with sentences that described marbles in bowls. In these sentences, the quantifiers only ranged over the amount of bowls. In the second type (quantifier) the amount of marbles were described by a scalar term (*many*, *some*, *a few*, *quite some*). For the first two filler types, the start word could either be *every*, *some* or *no*. The final type (numeral) is constructed of a numeral quantifying the amount of bowls, and a scalar term that quantifies the amount of marbles.

Table 3. Overview filler items

Reference	Example Sentence	Example Picture Description	Start word	Number of items

general	Every bowl contains marbles	(4) (5) (6)	Every/some/no	12
quantifier	Every bowl contains many marbles	(3) (4) (5)	Every/some/no	30
numeral	One bowl contains marbles	(2) (3) (2)	One/Two/Three[7]	42

For the pragmatic target items, it was not predictable what the *natural*/*unnatural* ratio would be. To balance this in the filler items, a third of the fillers was designed to lead to *natural* responses, a third to lead to *unnatural* responses, and a third to lead to mixed responses. The latter category was manipulated in the following way. The items that contained scalar terms were combined with images that would not clearly lead to a certain response. For each scalar term, we created a range for which we expected a *natural*, a *unnatural* or a mixed response. The bowls on the image were always filled with various amounts of marbles from one scale. The scale for *many*, for example, was as follows:

Table 4. Scale for *many*

Unnatural: 0-5	Mixed: 5-9	Natural: 9-12

Participants were exposed to the sentence three times, each time combined with an image that contained an amount of marbles that was associated with a different category of the range.

To control whether the experiment is suitable to trace SI's, a selection of the fillers is used for post-hoc control. This selection consists of underinformative sentences containing *some*. For example, a picture with three bowls filled with marbles was combined with (27). Recall that Bott and Noveck (2004) found mixed responses for underinformative sentences containing *some.* A second purpose of this subset was to create more items that lead to mixed responses.

(30) Some bowls contain marbles.

[7] The use of *three* is unpredictable. The use of *three* without a determinder might suggest the existence of a fourth bowl. This would have been avoided by using *the three*. A small delay for these filler items could be due to this. However, no relevant impact on the actual results is expected from this filler item.

It is known that SI's on *some* can be identified in an experimental setting (Bott and Noveck, 2004). If they cannot be found in this experiment, it might be related to the experimental design of our study. On the other hand, if SI's can be found for *some*, but not for the modifiers, it shows that the SI's for the modifiers differ from the one for *some*.

Table. 5 Overview control items

Reference	Example Sentence	Example Picture Description	Expected Response
pragmatic	Some bowls contain marbles	(3) (5) (6)	NAT/UNNAT
semantic	Some bowls contain marbles	(0) (6) (5)	NATURAL
false	Some bowls contain marbles	(0) (0) (0)	UNNATURAL

4.1.3 Participants

There were thirty eight participants in the experiment. They were recruited from Utrecht metropolitan area, the Netherlands. Participants' age ranged between 18 and 50 years old. Twenty-three of the participants were female and fifteen were male. All participants were university students and graduates and were native speakers of Dutch. Two participants were declared dyslexic. Twenty-eight participants participated voluntarily and were unpaid. Ten participants received a reward of five euros for their participation

4.1.4 Procedure

The experiment was conducted in a sound proof cabin. The only devices used were a computer screen and a response pad with three buttons. Participants were instructed to judge the sentence image combinations on naturalness by button presses. It was explicitly

Figure 2. The presentation of an example item on the screen.

mentioned that the grammaticality or stylistic form of the sentences was not important. The left/right response was counterbalanced between participants. For half of the participants, a left button press corresponded to *natural*, for the other half a left button press corresponded to *unnatural*. Before starting the experiment, participants were instructed how to hand the button box, to avoid a possible influence of switching buttons on the RT.

To optimize the reliability of the reaction times, participants had to parse the sentence by pressing a 'neutral' button in the middle of the device. The first thing to appear was always the picture, together with a text box filled with a plus sign (+) and two buttons with natural or unnatural. The plus sign was substituted by the first word of the sentence when the neutral button was pressed. For every button pressed a new sentence chunk would appear. Participants were told that they could press left or right the moment they saw a full stop. For the target items, the last chunk always consisted of the modifier, the numeral and the word *marbles*. By showing the relevant part at the final button press, we made sure participants could not start judging before finishing the sentence and thus ensured that the reaction time measured was the time between the final neutral press and the judgment press. By pressing the judgment button (left/right), participants got immediate feedback for their reaction on the screen by the disappearance of the answer option not chosen.

The items were presented in a pseudo-randomized order. The algorithm for the pseudo-randomization was based on the principle that the maximum number of appearances in a row of items containing a superlative modifier, a comparative modifier, or the start word every was three.

Participants were presented four example items to get used to the format of the experiment. The experiment took participants approximately 20 minutes. The personal data, including information on the age and gender of the participants was obtained through a questionnaire conducted after the experiment.

4.2 Results

Both the responses (*natural*/*unnatural*) and the RT's are analyzed. For the analyses, we distinguished three factors, namely modifier (2 levels), condition (3 levels) and numeral (7 levels). For modifier, we distinguished the factors superlative (*at least*) and comparative (*more than*). For condition we used the semantic condition (both semantically and pragmatically true), the false condition (both semantically and pragmatically false) and the pragmatic condition (semantically correct, and pragmatically false). For numeral, the levels consist of the numerals used in the experiment, which were all the natural numbers in the range from 3 to 9. In addition, we distinguished two random factors, namely subject and item. For the analyses, we used regression analyses to test the differences.

4.2.1 Data treatment

Before analyzing the data, two participants were removed from the dataset because of a remarkably high error rate (18% and 21% against an average of 3%). There were four observations of responses to fillers that were faster than 100 milliseconds. They were not relevant for the analysis, so no further steps are taken. There were no responses to target items that were quicker than 100 milliseconds. A log transformation was applied to the reaction time data. Errors were marked and removed from the RT analyses. For the target items, we marked all *unnatural* responses to items in the semantic condition, and all *natural* responses to the false condition as errors. For the fillers and control items, all natural responses to logically false items, and all unnatural responses to logically true items are were marked as errors.

4.2.2 Response Analysis

The distribution of the responses to the target items under different conditions can be found in the following table.

Table 6. Distribution of responses for both modifiers, including errors

Condition	Modifier	Response: natural	Response: unnatural
False	Comparative	19% (49)	81% (203)
False	Superlative	9% (23)	91% (229)
Pragmatic	Comparative	95% (239)	5% (13)
Pragmatic	Superlative	95% (239)	5% (13)
Semantic	Comparative	95% (240)	5% (12)
semantic	Superlative	93% (235)	7% (17)

Both in the semantic and pragmatic conditions, participants judged the statement as a natural description of the picture. There was no significant difference in response rates between the semantic and pragmatic conditions (β=-0,12, SE= 0,28, p=0,67) . The difference between the pragmatic and the false condition was significant (β=-4,77 SE= 0,24, p=<0,01).

To control for the possibility of deriving SI's under the experimental circumstances, a post-hoc analysis of a subset of the filler items is performed. This subset consists of the underinformative items containing *some*.

Table 7. Distribution of responses to the filler items containing *some*

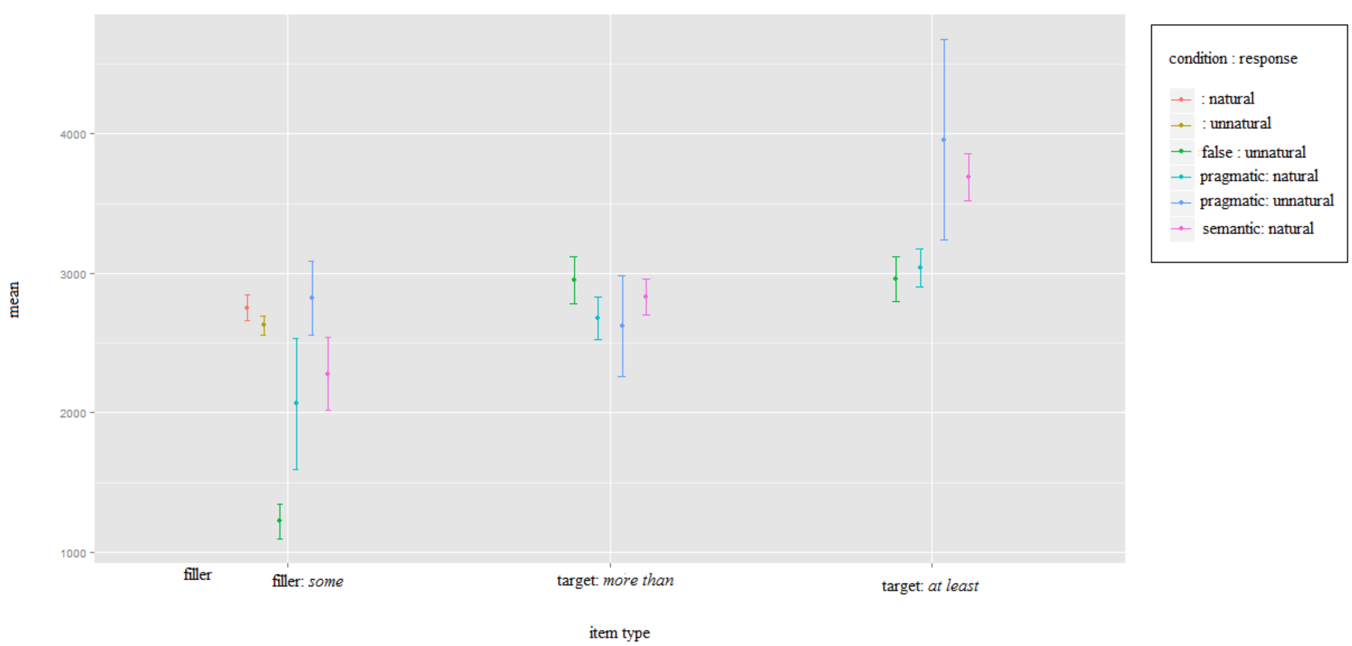
Discrepancy	Response: natural	Response: unnatural
False	5% (2)	95% (34)
Pragmatic	12,5% (9)	87% (63)
semantic	86% (31)	14% (5)

Table 6 shows that the experiment allowed for the rejection of pragmatically false items. In the case of the items containing *some*, most participants responded pragmatically, judging the statement as unnatural. A regression analysis shows that, despite the small amount of observations, the difference in responses between the pragmatic condition and the semantic condition is significant (β=3,77, SE= 0,60 p<0,01). No significant difference between the *pragmatic* and the *false* condition could be observed (β=-0,89, SE= 0,81, p=0,27). This shows that participants were very likely to give a pragmatic response to control items with *some*.

4.2.3 Reaction Time Analysis

The numerals used in the test sentences did not have an effect on the RT's. For the target items, the responses for the pragmatic condition were mostly semantic. There were not enough observations of pragmatic responses to compare the RT's of the semantic responses to the pragmatic ones. In the following table, the RT data is summarized.

Figure 3. Summary of the RT data



The graph shows that the RT's for *at least* are significantly longer than (β=-0,005, SE= 0,001, p<0,01). A linear mixed effects regression model shows an effect for *at least* (t=1,96) an likelihood ratio test shows this effect is significant (p<0,046). In addition, it shows that for the semantic condition, responses to the semantic condition are significantly slower than to both the natural responses for more than in the semantic condition (β=0,07, SE= 0,02, p<0,01) and the natural responses in the pragmatic condition (β=0,08, SE= 0,02, p<0,01). This is the same for more than. Responses for the semantic condition were significantly slower than responses for the pragmatic condition (β=0,08, SE= 0,02, p<0,01).

5. Discussing the Results

The results can be interpreted on two levels, the response data and the RT data. In this chapter, I will discuss these data and show how it relates to the theoretical framework of this thesis.

5.2 Scalar Implicatures on Modified Numerals embedded under Universal Nominal Quantifiers

The two theories discussed above both predict that a SI for *more than* can be generated when the modifier is embedded under a universal quantifier. More precisely, the UDM predicts that under a universal quantifier, SI's can be drawn from *more than*. Nothing specific is predicted for *at least*. The theory on non-monotonic alternatives predicts that both *more than* and *at least* allow for the derivation of SI's under universal nominal quantifiers.

The results however, clearly contradict these predictions. Participants did not derive implicatures for either modifier. The amount of observations for pragmatic responses was comparable to the amount of *unnatural* responses to the control items. There were no differences between *more than* and *at least*. In fact, the observations of natural and unnatural responses were exactly the same.

The conundrum of how to interpret these results is a difficult one. In the following section, I will discuss how the results can be interpreted

5.2.1 Some *generates SI's*

It is clear that the experimental design allowed for the derivation of SI's. A post-hoc analysis of the filler items containing *some*, shows that participants were very likely to respond pragmatically. In the experiment of this thesis, 87% of the responses to underinformative expressions with *some* were pragmatic. This is a high score, compared to results of other experiments in which *some* is tested in an underinformative condition. Bott and Noveck (2004) found that for their experiment, 61% of the responses was pragmatic. Banga et al. (2009) tested underinformative sentences with *sommige* the Dutch variant of *some* that was used in this experiment as well. They found that 75% of the participants responded pragmatically. Thus, the participants were more pragmatic than participants in the studies by Bott and Noveck (2004) and Banga et al. (2009)[8]. This shows that this experiment allowed for deriving SI's, which suggests that if modified numerals under universal quantifiers allow for SI's, they would have been traceable in the results.

In addition, note that the RT data for *some* is in line with Bott and Noveck's data(2004). The RT's for pragmatic responses were longer than RT's for semantic responses. This shows that the experimental design is reliable and likely to show effects for both pragmatic responses on RT's.

Let me return to the topic of this thesis: modified numerals under universal nominal quantifiers. With the high rate of pragmatic responses for *some*, it is remarkable that there were almost no pragmatic responses to the items containing modified numerals. The hypothesis is that SI's would be generated. In the following two sections, I will discuss two alternative hypotheses.

5.2.2 Implicature Strength

It is possible that SI's can be drawn from the sentences I tested in the experiment, but that participants did not. The question is whether an approach like this is compatible with the high rate of pragmatic responses for *some*. In this section, I will investigate the possibility of such an approach.

It is known that the actual derivation of a SI in general is dependent on various factors. For example, Doran *et al.* (2012) found a difference in the number of derived SI's between quantified statements and sentences containing adjectives. They performed an experiment in which participants had to judge underinformative sentences. They found that 32% of the quantified statements were rejected (which corresponds to a SI response), whereas the rejection rate for sentences with adjectives was only 17% of de adjectives. Scalar inferences were thus about twice as frequent for quantifiers as for adjectives. In other words, this shows that lexical scales are available to different degrees. In addition, van Tiel et al. (2014) argued that there are various factors factors that influence the availability of SI's, namely scalar distance, scalar boundedness, association strength, grammatical class, word frequencies and semantic relatedness.

For the interest of this thesis, this leaves us with speculations. *Some* is of a different grammatical class than modified numerals. This could possibly explain why there was a SI for *some* and not for the modified numerals. Word frequency cannot explain why an SI was drawn for *some* and not for the

[8] This difference can be related to different methodological factors. One difference between this experiment and the others is that in this experiment figures are used to create situations. This is a possible factor for the difference. Unfortunately, investigating this difference goes beyond the scope of this paper. The fact that it is an interesting observation that the experimental contexts influences the extent will which hearers reason pragmatically remains.

modified numerals. A quick google search provides 31 million hits for the Dutch counterpart of *some* (sommige) and 558 million hits for the Dutch counterpart of more than. The relevance lies in the fact that it is known that there are certain factors that influence the availability of SI's. In the experiment, we did not take such factors into account, so it is possible that the lack of SI's for the target items can be due to those factors. For this thesis, the most important thing is that there is clear evidence that the strength of SI's varies. The idea that if one SI (*some*) is available the other one (modified numeral) has to be available as well is too simplistic.

At this point, we could hypothesize that there is a difference in implicature strength between *some* and modified numerals. To support a hypothesis like that, it is most important to show, in a way that there exists a SI for modified numerals. Clearly, the experiment of this thesis did not support this hypothesis. An investigation of the availability of SI's for modified numerals that are embedded under universal nominal quantifiers is required.

Such an investigation could consist of data obtained in questionnaires or experiments. In the experiment of this thesis, participants were asked to evaluate with underinformative statements. This method did not show any SI's. For a follow up, there are multiple options. The first one is to ask participants to judge sentences on a Likert scale. Such an investigation would provide insight in the extent to which a sentence is natural for a hearer or not. The second option consists of a ranking task. If participants are asked to rank informative and underinformative sentences from 'most natural' to 'least natural', a SI, if present, will show up in the results.

If such methods do show a SI for superlative or comparative modified numerals under a universal nominal quantifier, these results can be linked to the theories discussed above. In that case, the difference in strength between *some* and modified numerals can be investigated as well.

5.2.3 Lack of Scalar Implicatures on Modified Numerals

The availability of deriving a SI for *more than* embedded under modified numerals under universal quantifiers is predicted by both Fox and Hackl (2006) and Mayr (2015). If no evidence can be found that SI's on *some* are stronger than on modified numerals, the results disconfirm these predictions.

First, Fox and Hackl (2006) argue that a dense scale causes a contradiction between the assertion and the potential SI in constructions containing modified numerals. This is not the case when they are embedded under a universal quantifier. In this way, the strengthened interpretation is derived. The results did not show SI's for that linguistic context. This disconfirms the UDM.

Mayr (2015) argues for an exhaustivity operator that only negates innocently excludable alternatives. The set of stronger alternatives for universal quantifiers embedding a modified numeral contains excludable alternatives – contrary to unembedded modified numerals – and therefore a SI is expected. The lack of SI's from modified numerals embedded under universal quantifiers disconfirms this theory.

5.3 The processing delay in the semantic condition

In addition to the responses, the RT's of those responses are measured. The motivation for measuring the RT's was to investigate whether there is a difference between pragmatic and semantic responses to SI's on modified numerals. In this section, I will discuss the results of the RT's results.

The hypothesis for the RT's in general was that responses to *at least* under all conditions would be slower than responses to *more than*. This hypothesis builds upon Geurts et al.(2010) and Cummins and Katsos (2010). The experiment showed a delay in general in *at least* compared to *more than*. It needs to be noted that there was no difference in the response times for negative responses. This is not something Geurts et al.(2010) or Cummins and Katsos (2010) showed. However, this is not surprising. For a negative response, participants do not need to fully process the meaning of the modifiers. More remarkable is the delay of positive responses to the semantic condition, compared to the pragmatic condition. Such a delay was not predicted, and is difficult to account for from the theoretical framework of this thesis. In this section, I will present a post-hoc analysis of this finding.

The finding that there occurs a delay for the semantic condition is surprising. From a theoretical point of view, one would expect the potential SI that is present in the pragmatic condition would cause a delay. This is not the case, participants are significantly faster in this condition, again this is applicable to both modifiers. Before deepening this, let us reconsider the conditions in the following table.

Table 7. The difference between the semantic and the pragmatic condition

Condition	Sentence	Picture
Semantic	Every bowl contains at least 5 marbles	(5)(6)(7)
Pragmatic	Every bowl contains at least 5 marbles	(6)(7)(8)

In first instance, the only difference between the items is the potential SI in the pragmatic condition. It is known that deriving this implicature is costly. Hence, any expected difference would be that the RT's for the pragmatic condition are delayed. The results show the opposite. The RT's for the semantic condition are longer. A second look at the target items illuminates a possible explanation. If no SI is calculated, which is the case, the only difference between the conditions is the distance from the numeral in the sentence – in the examples above 5 – to the number of marbles in the bowls.

I propose there is a correlation between the speed people process and the distance between the target numeral and the stimulus. Processing a modified numeral in combination with a visual stimulus that represents an x amount of objects, is faster if the distance between the numeral in the sentence and the x amount of objects is bigger. If participants have to judge whether a bowl contains more than 5 marbles, they will be faster if the bowl contains 100 marbles than if the bowl contains 6 marbles. This is the same for both *at least* and *more than*.

To investigate whether the distance between numeral and observed number underlies the delay in RT's, there are various options for further research. First, we expect a similar effect for other modified numerals. A similar experiment could be conducted in which modifiers like *from* and *minimally* are included. In addition, upperbound numerals can be included as well, for example *less than* and *at most*. Further, in a follow up study different sizes of distance can be compared. In the experiment of this thesis, we only included distances from 0 to 2. We could only compare a small distance (0 for superlative, 1 for comparative) with a less small distance (1 for superlative, 2 for comparative).

In the field of the semantics of modified numerals, the importance of this finding lies in the methodological value. Future researchers need to consider the effect of distance between a stimulus numeral and a numeral in a sentence when conducting experiments comparable to the one described in this thesis.

5.4 Final conclusion

In this thesis, I investigated the availability of SI's under modified numerals when embedded under a universal nominal quantifier. The goal was to contribute to this topic by providing quantitative data and an experiment was conducted. The conclusions are as follows.

First, the experiment showed that the distance from the numeral that is modified to the number of observed items plays a role in the processing speed. If the distance is larger, the processing speed increases. This is an interesting finding, and future research has to show whether this is applicable to all modifiers, and perhaps even other scales. In addition, it is a valuable methodological issue that future researchers need to take into account when carrying out experiments of a comparable design.

Second, the experiment in which underinformative sentences were tested showed that no SI's were derived from modified numerals in under a universal nominal quantifier. I argue that there are two possibilities. The first option is that the lack of SI's in the results of the experiment is due to implicature strength. In this case, further research is required to show that SI's can be drawn from modified numerals in this linguistic context. If this is the case, this experiment has shown that SI's on modified numerals under universal quantifiers are not as strong as SI's on the scalar term *some*, which is in itself a valuable finding. The second option is that no SI's available for modified numerals when embedded under a universal nominal quantifier. In this case, this is a gap in both the UDM and the theory on non-monotonic alternatives.

7. References

Banga A, Heutinck I, Berends S M, Hendriks P . 'Some implicatures reveal semantic differences'. In: Botma B, van Kampen J, editors. Linguistics in the Netherlands 2009. Amsterdam: John Benjamins; 2009. p. 1-13.

Bott, L., and Noveck, I. (2004). Some Utterances are underinformative: The onset and time course of scalar inferences. Journal of Memory and Language 51:437-457.

Büring, D. (2007). The least „at least" can do. In C. B. Chang and H. J. Haynie (eds.), Proceedings of WCCFL 26. Somerville, MA: Cascadilla Press. 114-120.

Cummins, C., & Katsos, N. (2010). Comparative and Superlative Quantifiers: Pragmatic Effects of Comparison Type. Journal of Semantics, 27: 271-305.

Doran, R., Baker, R. E., McNabb, Y., Larson, M., & Ward, G. (2009). On the non-unified nature of scalar implicature: an empirical investigation. International Review of Pragmatics, 1(2), 211-248.

Geurts, B., Katsos, N., Cummins, C., Moons, J., & Noordman, L. (2010). Scalar Quantifiers: Logic, Acquisition, and Processing. Language and Cognitive Processes, 25(1): 130-48.

Geurts, B. and Nouwen, R. (2007). At least et al.: the semantics of scalar modifiers. Language, 83:533-559.

Grice, H. P. (1975). Logic and Conversation. In P. Cole and J.L. Morgan (eds.), Syntax and Semantics, Vol. 3. New York: Academic Press. 41-58.

Groenendijk. G., Stokhof, M. (1984): Studies on the Semantics of Questions and the Pragmatics of Answers. PhD dissertation, University of Amsterdam.

Fox, D., & Hackl, M. (2006). The universal density of measurement. Linguistics and Philosophy, 29(5), 537-586.Hackl, M. (2000). Comparative quantifiers, Doctoral dissertation, MIT.

Hackl, M (2000): Comparative Quantifiers. PhD dissertation MIT, MITWPL, Cambridge, MA.

Horn, L. (1972). On the semantic properties of logical operators in English. UCLA dissertation, distributed by Indiana University Linguistics Club, 1976.

Krifka, M. (1999). At least some determiners aren’t determiners. In K. Turner (ed.), The Semantics/Pragmatics Interface from Different Points of View, Current Research in the Semantics/Pragmatics Interface Vol. 1. Oxford: Elsevier. 257-92.

Lea, R.B. (1990). Predicting Propositional Logic Inferences in Text Comprehension. Journal of Memory and Language 29, 361-387.

Levinson, Stephen C. Presumptive meanings: The theory of generalized conversational implicature. Cambridge, MA: MIT press, 2000.

Mayr, C. (2015) Implicatures of modified numerals. In I. Caponigro and C. Cecchetto (ed.) From Grammar to Meaning, The Spontaneous Logicality of Language. Cambridge University Press. 139-171.

McNabb, Y., & Penka, D. (2013). The interpretation of superlative modifiers and deontic modals: An experimental investigation. Presentation at Sinn und Bedeutung.

Musolino, J. (2004). The semantics and acquisition of number words: integrating linguistic and developmental perspectives. Cognition, 93:1-41.

Nouwen, R. (2008). Upperbounded no more: the exhaustive interpretation of non-strict comparison. Natural Language Semantics, 16(4): 271-95.

Nouwen, R. (2010). Two kinds of modified numerals. Semantics and Pragmatics, 3(3): 1-41.