

Testing the nature of variation effects with modified numerals*

Stavroula Alexandropoulou — *Utrecht University*

Abstract. Variation effects, which are akin to free choice effects, are triggered when a numeral is modified by the lower-bound superlative modifier *at least* and appears in certain embedded contexts. The dominant take on the derivation of those effects is a pragmatic view whereby they arise via an implicature-generating mechanism. In this paper, I present results from two experiments that tested the availability of variation inferences with lower-bound class B numeral modifiers in the scope of a universal quantifier as well as whether those inferences have a semantic or a pragmatic strength. I show that variation effects i) do arise, and ii) are pragmatic inferences, as predicted by the pragmatic view. Moreover, the findings of this study are compatible with an alternative-introducing semantics for *at least* à la Büring (2008) and Coppock and Brochhagen (2013).

Keywords: modified numerals, variation effects, semantic/pragmatic inferences, experimental semantics/pragmatics

1. Introduction

Most of the literature on the semantics and pragmatics of superlative numeral modifiers has been dealing with the well-established ignorance or speaker insecurity (SI) inferences (Büring 2008) they give rise to (e.g., Geurts and Nouwen 2007; Cummins and Katsos 2010; Nouwen 2010; Schwarz 2013). Interestingly, those inferences are principally obviated when such modifiers appear in certain embedded contexts, and other inferences are triggered instead. The present paper is concerned with those latter, less studied and established inferences.

Take example (1) below, where the superlative modifier *at least* interacts with the universal quantifier *every*. The most preferred reading conveyed by (1) consists of the basic meaning of (1), that is, ‘for each laptop it is the case that the number of GB of memory is ≥ 2 ’ plus the following meaning component: ‘the number of GB varies with respect to laptops’; for instance, laptop A has 2GB *and* laptop B has 4GB *and* laptop C has 4GB *and* laptop D has 8GB, et cetera. The latter, additional meaning component constitutes an example of those inferences superseding SI inferences in the presence of an operator, such as the universal nominal quantifier in the example below.

(1) Every laptop we sell has at least 2GB of memory. (adapted from Nouwen 2015)

Büring (2008) has attributed to the whole conveyed reading the characterisation ‘authoritative’ to

*The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. 313502. I would also like to thank the SuB19 audience, Rick Nouwen, Jakub Dotlačil, Maartje Schulpen, Dominique Blok, Yaron McNabb, Sophia Manika, and Mattis van den Bergh for their helpful feedback, as well as all my participants.

contrast it with the least preferred, subsiding SI reading, which in (1) would be ‘there is a certain n of GB of memory each laptop has and the speaker is not sure whether this is 2 or more’. The authoritative reading arises when the *at least* phrase scopes below the universal quantifier, and so it is, when *at least* interacts with other operators, such as universal modals, plurals, generics (Nouwen 2015); see an example of a universal modal in (2).

(2) Sophia has to write at least fifteen pages.

As in (1), the arising inference says that ‘the number of pages can vary’, e.g., Sophia can write 15 pp, *and* she can write 17 pp, *and* she can write 20 pp, etc., which we can be represented as follows:

w_1 : 15 pages

w_2 : 17 pages

w_3 : 17 pages

w_4 : 20 pages

...

where w stands for a deontically accessible world

In both examples (1) and (2), we observe that there is a ‘variation’ output from the interaction of *at least* with the respective present operator. Thus, in (1) we get a variety or range of pairs of laptops and numbers of GB, and of pairs of worlds and numbers of pages in (2). This is the result of the distribution of relevant numbers of GB over individual laptops in the former case and of relevant numbers of pages over deontically accessible worlds in the latter case. That latter case is known as the ‘distribution requirement’ (Kratzer and Shimoyama 2002). Note also that, independently of what the embedding operator is, the output range is expressed as a conjunction of entities/situations (notice the italicised ‘and’ in the paraphrases of (1) and (2)); for instance, in (2), the output is a conjunction of permissions or choices, depending on the perspective.

The variation effects of (1) and (2) including the conjunctive output resemble the variation output of free choice (fc) effects with disjunction (Kamp 1973; Zimmermann 2000; Nickel 2010, *inter alia*). See (3) for an illustration, where *or* interacts with a universal nominal quantifier:¹

(3) Every student listens to Pepper fm or to En Lefko fm.

(3) conveys the fc reading ‘some students listen to Pepper fm and some students listen to En Lefko fm’; or in other words, (3) is true in the following variation scenario:

¹Obviously, the effect in (3) is not a genuine fc effect, since there is no choice involved. A genuine fc effect arises when we are dealing with modal quantifiers. Similarly, the variation effect in (1) corresponds to the so-called ‘modal variation’ effect in the modal domain (cf. Alonso-Ovalle and Menendez-Benito 2010), which we have in example (2).

Student A listens to Pepper fm
 Student B listens to Pepper fm
 Student C listens to En Lefko fm
 Student D listens to En Lefko fm
 ...

However, the variation effect with *at least* is weaker than that in free choice, as Nouwen (2015) points out and shows by means of the example in (4):

- (4) a. **Context:** *Password policy*: For security reasons, the system will not accept passwords that are shorter than 6 characters. Moreover, it cannot handle passwords that are longer than 10 characters.
 b. Passwords have to be at least 6 characters long.

(4-b) is true given (4-a). If the variation effect in (4-b) were the same as that of free choice, (4-b) should not be true, because it would entail that a password of *any* number of characters greater than five would be an acceptable one. Hence, this suggests that the variation effect with *at least* seems to merely say that — after meeting the requirement of $n_{char.} > 5$ — ‘there is no specific n such that passwords need to be exactly n characters long’. For this reason I will not make use of the term ‘free choice’ or ‘free choice-like’ to refer to this — weaker/partial, as opposed to universal — effect I am studying here, but I will stick to the descriptive and plain term ‘variation effect’.

When dealing with such little studied effects, the first questions come to mind are the following: i) Do variation effects with superlatively modified numerals exist? ii) If so, what is the strength of those effects? Do they have the strength of a semantic or that of a pragmatic inference? These are the very questions the present paper seeks to answer experimentally, zooming in on variation effects with lower-bound numeral modifiers, such as *at least* and *n or more*.

In the next section I present the pragmatic view on the variation effects with superlatively modified numerals, in which Coppock and Brochhagen’s (2013) analysis features. In Section 3, I report on the experiments I conducted in order to test the predictions that follow from the pragmatic view with respect to the research questions specified above. The final section concludes.

2. Pragmatic view: Variation effects as implicatures

The questions posed in the previous section have dominantly been tackled from a pragmatic perspective. More specifically, according to this perspective, variation effects arise via an implicature-generating mechanism. However, within that perspective there are two kinds of accounts for the generation of variation effects as implicatures. Both kinds are Gricean, but they differ in how alternatives come about. More precisely, there are those accounts that derive the variation effects

via Gricean reasoning and crucially start off with an alternative-introducing semantics for *at least*, and those that execute the standard Gricean recipe for scalar implicatures assuming lexically determined sets of alternatives.

As to the latter camp, let us briefly see how the effect we are after is derived via a scalar implicature-generating mechanism. Consider example (1) from the Introduction; assuming the scale of natural numbers excluding 0 (i.e., \mathbb{N}^*) as the relevant scale and applying the scalar implicature reasoning, we get the implicature ‘not every laptop we sell has 3 or more GB of memory’. This in combination with the assertion suggests that some laptop in that shop has exactly 2GB of memory. That is, some laptop has (or some laptops have) exactly 2GB of memory and some laptops have more than 2GB. It is further entailed that ‘not all laptops of the shop have the same amount of memory’, which is exactly what the variation effect suggests. Mayr’s (2013) and Schwarz’s (2013) accounts fit in the pragmatic camp in question, and both stipulate distinct scale mates for the superlative modifier and the numeral in order to derive the scalar implicature triggered when *at least* is in an embedded environment. I will not go into the details of those accounts, but I will rather turn to the other pragmatic camp, whose more specific predictions will be tested in Experiment 2.

Proponents of the first camp are Coppock and Brochhagen (2013) and Büring (2008), and among the modified numeral literature the former have elaborately dealt with variation readings. Their account of superlative modifiers takes inspiration from Büring’s (2008) prior idea that superlative modifiers are disjunctions at some level of description. In what follows I flesh out this pragmatic view by the illustration of Coppock and Brochhagen’s (2013) more detailed account.

2.1. Coppock and Brochhagen (2013)

Coppock and Brochhagen (2013) assume an alternative-introducing semantics for superlative modifiers within an inquisitive semantics framework, according to which superlative modifiers denote as many possibilities as the alternatives under consideration (see Ciardelli, Groenendijk, and Roelofsen 2012, for the relevant terminology). Below you see a straightforward illustration of that idea.

- (5) $[[at\ least]]^s(n)(A)(B) = \{ |A \cap B| = n, |A \cap B| = n+1, |A \cap B| = n+2, \dots \}$, where s stands for *state*, that is, the current discourse context, and $n \in \mathbb{N}^*$.²

(5) shows that *at least* creates sets of propositions that encompass the alternatives being ranked as high or higher on a pragmatically defined scale. In example (6), the relevant set of alternatives would be {Magda called 5 times, Magda called 6 times, Magda called 7 times, ...}, and this constitutes the core of the semantics Coppock and Brochhagen (2013) attribute to (6).

²Here I am zooming in on the cases where *at least* takes a numeral as an argument.

(6) Magda called at least five times.

Let us now turn to their account of variation effects, starting with a small introduction. The implication in (7) summarises the observation of arriving at a(ny) variation effect with conjunctive nature from an embedded alternative-introducing or disjunctive expression. More precisely, in the antecedent of the implication we have the basic, asserted, meaning with the alternatives/disjuncts introduced by the relevant expression, while the consequent of (7) consists of the additional meaning component of the conjoined permissible options in the case of a universal modal and of the conjoined existentially bound alternatives in the case of a universal nominal quantifier.

(7) $\Box\forall (\alpha \vee \beta) \rightarrow \Diamond\exists \alpha\exists \wedge \Diamond\exists \beta$ (Note that I introduce this notation for the nominal quantifiers in order to demonstrate the observed similarity independently of the operator the alternative-based expression is embedded under.)

But how do we get from the left to the right part of the implication? In other words, how is the variation effect in the consequent derived from the antecedent containing the relevant alternatives? In order to answer the above question, concerning the variation effects with superlatively modified numerals in the scope of universal quantification, and starting with an alternative-based semantics, Coppock and Brochhagen (2013) execute Kratzer and Shimoyama's (2002) Gricean recipe. This recipe was devised to account for the fc effects of the German epistemic indefinite *irgendein* in the scope of a universal modal; see (8), taken from Kratzer and Shimoyama (2002).

(8) *Du musst dir irgendeins von diesen beiden Büchen leihen.*
 you must you.DAT irgend-one of those two books borrow
 'You must borrow one of those two books.'
 \rightsquigarrow You are allowed to borrow book A *and* you are allowed to borrow book B.

The translation of (8) spells out the asserted content of (8), and below that you see the fc inference that is triggered; that is, the referent of *Du* is free to choose between book A and book B. Crucially, Kratzer and Shimoyama (2002) too assume that *irgendein* is an alternative-introducing, here disjunctive, expression, with book A and book B each being a relevant stronger alternative in the discourse context of (8). They derive the fc reading by the following Gricean reasoning: Why did the speaker in (8) choose to use *irgendeins*, with an alternative-introducing/disjunctive semantics, that is, picking the widest set of relevant alternatives you see in (9-a), rather than a more specific and stronger alternative, such as $A = \{\text{You borrow book A}\}$? Because it is either the case that $\Box A$ is false or that it is true, but its exhaustivity inference in (9-b) is false. Then it follows that $\Box A \Rightarrow \Box B$. Applying the same reasoning to the remaining alternative in ALT, i.e., $B = \{\text{You borrow book B}\}$, we arrive at $\Box B \Rightarrow \Box A$. Hence, we get the following: $\Box A \Leftrightarrow \Box B$. Combining this equivalence with the asserted meaning, as notated in (7) by the left part, it follows that $\Diamond A \wedge \Diamond B$. As is obvious,

we have generated the consequent of (7), which corresponds to the fc inference of (8), that is, ‘you are allowed to borrow book A and you are allowed to borrow book B’.

- (9) a. $ALT = \{\text{You borrow book A, You borrow book B}\}$
 b. $\neg \Box B = \{\text{You borrow book B}\}$ *exhaustivity inference of A*

As already said, Coppock and Brochhagen (2013) apply the same rationale in order to derive the variation effect of sentences like (1), repeated below as (10).

- (10) Every laptop we sell has at least 2GB of memory.

Let us see how. Why did the speaker of (10) pick an expression with an alternative-based semantics, namely, one denoting the set in (11-a) instead of a stronger alternative, e.g., $A = \text{Laptop } x \text{ has 2GB}$? Via the same reasoning as above we arrive at $\forall -A \Leftrightarrow \forall -B$ (notation $\forall -A$ is introduced to abbreviate embedding under \forall). This in combination with the basic semantics of (10) in (11-b) derives the live options of memory of laptops sold by the store of the discourse context of (10), represented as $\exists -A \wedge \exists -B$. Notice that this corresponds to the right part of the implication in (7).

- (11) a. $ALT = \{\text{Laptop } x \text{ has 2GB, Laptop } x \text{ has 4GB}\}^3$
 b. $\forall -(A \vee B)$ (alternative-based expression in the scope of \forall , cf. left part of (7))

Before ending this section, I would like to briefly mention that Büring (2008), who proposed that *at least n* is a disjunction, being interpreted as ‘exactly *n* or more than *n*’ (also adopted by Cummins and Katsos 2010), in order to derive the variation effect (or the authoritative reading in his own terms) of *at least n* embedded under a universal modal, adopts a very similar scheme to that in (7), which he actually borrows from Klinedinst (2007).

To conclude Section 2, both Coppock and Brochhagen (2013) and Büring (2008), as well as the pragmatic camp of scalar implicatures, provide accounts for the derivation of variation effects with *at least* and they all put forth a pragmatic analysis to derive those effects; for this reason they are subsumed under the term ‘pragmatic view’. In what follows, I present an experimental investigation of the predictions stemming from the pragmatic view as regards the questions of the availability and of the semantic/pragmatic strength of the effects in question. I additionally test

³Similarly to Kratzer and Shimoyama’s (2002) example above, I am limiting the set of alternatives to two. Both in Kratzer and Shimoyama (2002) and in Coppock and Brochhagen (2013) it is implied that a straightforward generalisation in the case of three or more alternatives generates the output of $\diamond A \wedge \diamond B \wedge \diamond C \wedge \dots$, and that of $\exists -A \wedge \exists -B \wedge \exists -C \wedge \dots$, respectively. However, further clarification would be in order, as the details of this generalisation are not spelled out by the aforementioned authors.

predictions of Büring’s (2008) and Coppock and Brochhagen’s (2013) pragmatic view with respect to an alternative-based status for *at least*.

3. Experiments: Test the availability and strength of variation effects with *at least*

3.1. Predictions

According to the pragmatic view, it is predicted that variation effects with *at least* do arise and, more specifically, they are argued to arise via pragmatic reasoning; thus, they are predicted to be as strong as pragmatic inferences. To test these predictions, I conducted two offline experiments, both in the form of a questionnaire eliciting people’s judgements.

3.2. Experiment 1

3.2.1. Methods

Experiment 1 was a paper and pencil questionnaire consisting of short dialogues between a person A and a person B. Person A made a statement and person B followed up with a question. Participants were tasked to judge how reasonable B’s follow up was given A’s statement. They did so by giving ratings on a –5 to +5 Likert scale, where –5 is ‘completely unreasonable’ and +5 is ‘completely reasonable’. This scale has been inspired by a similar one introduced by Cummins and Katsos (2010). The rationale for using such a scale is that semantic contradictions are expected to score on the left side of the scale and close to –5, and semantically and pragmatically well-formed items are expected to score on the other side of the scale and close to +5. Pragmatic infelicities should be rated higher than the semantic contradictions, but still lower than the semantically and pragmatically well-formed items. This is arguably a way to draw a distinction between semantic contradiction and pragmatic infelicity, as Cummins and Katsos’s (2010) relevant results reveal. Also, if a pragmatic inference arises less reliably or is by no means obligatory, a large variation of scores is expected. A similar scale has also been employed by McNabb and Penka (2014), suggesting a trustworthy method.

Below you find an example of my experimental sentences, in which *at least* interacted with the universal nominal quantifiers *elk* ‘every’ or *iedereen* ‘everybody’:

- (12) **A:** *Volgens een steekproef zitten er in elk zakje minstens tweeëntwintig*
 according.to a random-sample sit there in every bag at.least 22
dropjes.
 licorice-candies
 ‘According to a random sample every bag contains at least 22 licorice candies.’

- B:** *Zitten er evenveel dropjes in elk zakje?*
 sit there as.much licorice-candies in every bag
 ‘Do they all contain the same number of licorice candies?’

The asserted meaning of A’s utterance is ‘for every sack the n of licorice candies is ≥ 22 ’. B’s follow up constitutes a reasonable reaction to A when merely considering A’s assertion. Speaker A can additionally convey that ‘there is no specific n such that all sacks contain exactly that many licorice candies’ (variation reading). B’s question targets this latter meaning and it serves as an unreasonable and infelicitous follow up to A given that meaning, because it asks whether the exact opposite is the case. To see whether this infelicity has the strength of a contradiction or that of a pragmatic infelicity, I included semantically contradictory dialogues as well as pragmatically ill-formed/infelicitous dialogues as control items, also used as baselines. B and B’ in (13) are two respective examples. All those control items involved disjunction in interaction with the universal quantifiers *elk* ‘every’ or *iedereen* ‘everybody’, similarly to the experimental items.

- (13) **A:** *Bij de lunch heeft iedereen op het werk een salade of een kom soep besteld.*
 at the lunch has everyone at the work a salad or a bowl soup ordered
 ‘Everybody at work ordered salad or soup for lunch.’
- B:** *Was er iemand van hen die geen van beide heeft besteld?*
 was there anyone of them who none of the.two has ordered?
 ‘Did anyone order neither of those?’ *semantic contradictions*
- B’:** *Was er iemand van hen die een kom soep heeft besteld?*
 was there anyone of them who a bowl soup has ordered
 ‘Did any of them order soup?’ *pragmatic infelicities*

A’s assertion states that one of the two disjuncts is true (basic meaning of disjunction), and is true when everyone had a salad and when everyone had soup. B’s follow up in semantic contradictions is completely unreasonable, because it prompts for a proposition that contradicts A’s assertion. Speaker A can further convey the fc reading that ‘some people ordered a salad and some people ordered a soup’ (see Nickel 2010). B’s follow up in pragmatic infelicities is unreasonable as well, because it questions what the fc reading of A’s utterance states.

Semantically and pragmatically well-formed/felicitous control items were included too, being also used as a baseline. The interacting DPs in these controls were varied (that is, *every* NP/proper name * modified numeral/disjunction/definite description) to achieve distraction and to have also well-formed/felicitous items involving the interactions *every* * modified numeral/disjunction. (14) illustrates a semantically and pragmatically well-formed dialogue, where A’s statement contains an *every* NP (*elke student*) and a definite description (*het Van Gogh Museum*).

- (14) **A:** *Vorige maand heeft elke student kunstgeschiedenis het Van Gogh Museum bezocht.*
 previous month has every student history.of.art the Van Gogh museum
 visited
 ‘Last month every history of art student visited the Van Gogh museum.’
- B:** *Vonden ze het interessant?*
 found they it interesting
 ‘Did they find it interesting?’

B’s question is a natural and felicitous, thus, reasonable, follow up to A’s utterance. Juxtaposing the experimental items with those controls will tell us to what extent the infelicity described above for the experimental item (12) actually arises. In other words, a comparison of our experimental items with the semantically and pragmatically well-formed/felicitous control items will indicate whether variation readings are at all available with *at least*. If it turns out that variation readings are indeed available, the comparison in question could furthermore manifest how robust those readings are. Participants were provided both with an example of a semantically contradictory case as well as with a semantically and pragmatically well-formed example in the instructions of the questionnaire. Those examples did not involve any of the configurations used in the test items or the semantically contradictory and the pragmatically ill-formed/infelicitous control items.

A version of the experimental items with the numeral modifier *meer dan* ‘more than’ was also tested, but I am not discussing here the relevant results. In the experiment there was one factor, CONDITION: AT LEAST, MORE THAN, semantic contradictions, pragmatic infelicities, and semantically and pragmatically well-formed control items. The task included six experimental items (two conditions: AT LEAST and MORE THAN conditions) and six control items with disjunction (two conditions: condition of semantic contradictions and condition of pragmatic infelicities). Each type of items was divided into two lists, so that each participant saw each item only in one condition. 14 semantically and pragmatically well-formed dialogues appeared in each of the lists, also serving as distractors. For the purpose of counterbalancing the total design, two filler items expected to score very low (involving neither numeral modifiers nor disjunction) were added (number of stimuli = 28). Each of the formed lists of items appeared in two different orders, yielding four lists to test in total. A total of 27 subjects (20 Female; Mean age: 19.3; Age range: 17–23), all bachelor students at Utrecht University and native speakers of Dutch, filled in the questionnaire voluntarily and were naive as to the purpose of the study. Overall 756 observations were obtained.

3.2.2. Results

The reasonability judgements, or else scores, obtained were ordinal data, so they were analysed with ordered probit models using the `ordinal` package (Christensen 2013) in R. AT LEAST was the reference level of the factor CONDITION and was compared to MORE THAN and all three types

of control items. The model also included intercept random effects for subjects and items.

Condition AT LEAST was rated significantly lower than the semantically and pragmatically well-formed control items ($\beta = -1.776, SE = .174, p < .0001$). It further got scores significantly higher than the semantic contradictions ($\beta = 1.808, SE = .225, p < .0001$) as well as than the pragmatic infelicities ($\beta = 1.284, SE = .212, p < .0001$). These come as a statistical confirmation of the relevant differences one can see on the plot (Figure 1) between AT LEAST, on the one hand, and each of the control conditions on the other hand. Further analyses showed that the difference between semantic contradictions and pragmatic infelicities as well as the difference of the former from the well-formed control items were statistically significant ($\beta = -.508, SE = .191, p = .008$; $\beta = -3.408, SE = .236, p < .0001$, respectively). So was the difference of the pragmatic infelicities from the well-formed control items ($\beta = -3, SE = .242, p < .0001$).

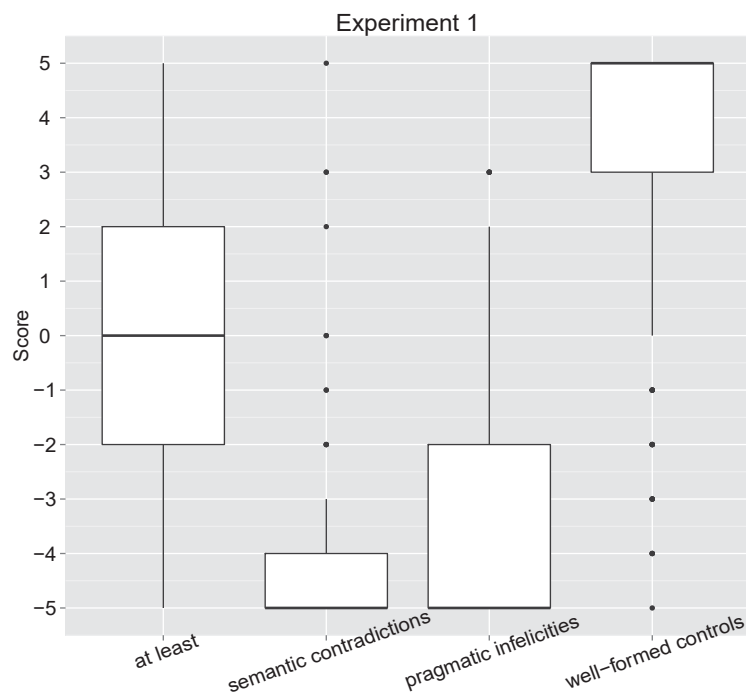


Figure 1: Boxplots of scores per condition. Experiment 1.

3.2.3. Discussion

The significant difference between the AT LEAST items and the well-formed control items indicates that subjects found that B's questions were not completely reasonable follow ups to A's statements, i.e., that something in B's questions gave rise to infelicity. As specified in Section 3.2.1, this would be so in case A's statements were understood as conveying a variation reading. Thus, I conclude

that variation effects were attested in the AT LEAST items. This lends support to the first prediction according to the pragmatic view I am testing, namely, that variation effects arise when *at least* interacts with the universal nominal quantifier. The question that now follows is if we can tell whether these effects have the strength of an entailment or that of a pragmatic inference. The difference of the AT LEAST condition from the semantic contradictions reported above shows that subjects did not rate the former as bad as the latter, that is, they did not treat them as contradictions. From that it is inferred that the attested variation effects with *at least* are not as strong as entailments, pointing to them having the strength of a pragmatic inference. Hence, the second prediction that variation effects are derived via a pragmatic mechanism is also borne out.

I additionally found that AT LEAST items were rated significantly higher than the pragmatic infelicities. One could speculate that this difference suggests that variation inferences with *at least* are weaker inferences than the fc effects with disjunction found in the pragmatic infelicities; that is, the former are generated less reliably than the latter.⁴ Or rather one could interpret this difference as follows: the variation inference-generating mechanism with *at least* is different from the fc inference-generating mechanism with disjunction. This would speak against Büring's (2008) version of the pragmatic view that wants *at least* and disjunction to make use of the same mechanism in the derivation of the respective inferences. Such a conclusion would however be a remarkably weak one given the recent findings in the experimental literature on scalar inferences, cf. van Tiel, van Miltenburg, Zevakhina, and Geurts (2014).⁵ More specifically, van Tiel et al. (2014), building on Doran and colleagues' (2009) prior work, showed that different scalar expressions, such as quantity expressions (e.g., *some*), adjectives like *warm*, *good*, *big*, adverbs like *sometimes*, *possibly*, verbs like *like*, *might*, *try*, diverge in the rate at which they give rise to scalar implicatures (via the standard Gricean reasoning). Thus, this could not permit the conclusion that the inference-generating mechanism for *at least* and for disjunction is different.

A particularly plausible interpretation of the significant difference found between the AT LEAST condition and the pragmatic infelicities was suggested to me by an anonymous reviewer and is as follows.⁶ The difference between the conditions in question could very well signal that ignorance readings have kicked in. Experimental items and pragmatically ill-formed control items involve an interaction between *at least* and disjunction, respectively, on the one hand, and the universal quantifier *elk* 'every' or *iedereen* 'everyone', on the other hand. As mentioned already in the Introduction, the most preferred output of such an interaction is the reading derived when *at least*, or disjunction, takes narrow scope with respect to the universal quantifier. As Büring (2008) and Nouwen (2015) have pointed out, the least preferred reading is the one where *at least* and disjunction scope over the operator they interact with.⁷ As already said, the resulting readings have been dubbed ignorance or speaker insecurity readings (Büring 2008). See below the two possible

⁴See a similar finding and interpretation in Cummins and Katsos (2010) with respect to the intermediate position of ignorance effects with modified numerals between their pragmatic infelicities and well-formed control items.

⁵Thanks to an anonymous reviewer for pointing this out.

⁶Thanks to this reviewer for the very useful and valuable comment.

⁷See McNabb and Penka (2014) for experimental work on those readings, where *at least* interacts with modals.

readings for the experimental item in (12) and the pragmatically ill-formed control item in (13).

- (15) **A:** According to a random sample every bag contains at least 22 licorice candies.
B: Do they all contain the same number of licorice candies? AT LEAST
 ~> There is no specific n such that all bags contain exactly that many liquorice candies. *variation reading*
 ~> Every bag contains n candies, and as far as the speaker knows n could be 22 or more. *SI reading*
- (16) **A:** Everybody at work ordered a salad or a soup for lunch.
B: Did any of them order a soup? DISJUNCTION
 ~> Some people ordered a salad and some people ordered a soup. *fc reading*
 ~> Everybody ordered the same thing, and the speaker does not know whether that was salad or soup. *SI reading*

Assuming that both readings are possible for each statement A, there is a difference in the expected judgement of question B in the two conditions, even if the respective inferences are of the same strength. In (16), question B is expected to be judged as an unreasonable follow up to A, on either reading. Take the *fc* reading; as we have already seen, in that case question B is infelicitous because, simply put, what it enquires about has just been stated by A. Now take the *SI* reading; again question B is infelicitous, because it asks the ignorant speaker A for knowledge that A lacks; that is, whether the one, and the same, thing everyone ordered for lunch — which is the only piece of knowledge A has — was a salad or a soup. Thus, regardless of what reading subjects get for A in the pragmatically infelicitous controls, they are expected to receive low scores. This is not the case for the experimental items with *at least*, because question B is more easily and straightforwardly interpreted as probing whether A meant to convey the variation reading or the *SI* reading. Notice, however, the *according to* phrase in (15); this arguably weakens the availability of an *SI* reading. Yet, this item did not score very low (mean = -.267), contrary to one's expectation that the *SI* reading weakening by the *according to* phrase would make the variation reading more salient, and thus, the item more susceptible to score low. More importantly, not all experimental items were constructed in a similar way, that is, containing such prepositional phrases or other phrases with similar function. In addition, in the experimental items, besides a general avoidance of round numbers to be modified by *at least*, which being a sign of impreciseness (cf. Krifka 2009, *inter alia*) facilitate *SI* readings, no systematic and sophisticated means of controlling for the interference of *SI* readings were employed. Consequently, it is highly plausible that the inflow of *SI* readings by the interpretive strategy recommended by the reviewer caused the relevant items to receive higher scores than expected, and thus, also higher than the pragmatically infelicitous controls with disjunction. This perfectly explains the difference I found between the *AT LEAST* items and the pragmatic infelicities and, beyond doubt, leads to conducting Experiment 2, the set up of which was modified in such a way as to weaken the *SI* readings' interference and have

participants focus on the variation readings.

3.3. Experiment 2

Experiment 2 was a follow up study aiming at showing whether the SI readings constituted a potential factor in driving the AT LEAST scores high in Experiment 1, getting more clear data.

3.3.1. Methods

Experiment 2 consisted of lists that were randomly distributed and filled in mostly online. A few data were collected afterwards in a paper and pencil fashion, in order to balance the number of observations across lists. The lists of this experiment were created on SurveyMonkey (www.surveymonkey.com), where links to each list were generated.

Subjects were instructed to read short dialogues between a researcher and an interviewer. The researcher makes a statement about the findings of a recent successful research s/he was involved in and the interviewer asks a question about them. This modification of the instructions was inspired by the instructions used in relevant work in progress by J. Dotlačil and R. Nouwen (p.c.), and already constitutes a deviation from Experiment 1, serving to hinder the interference of SI readings. To be more specific, the fact that the instructions made clear that the researcher had direct involvement in the research whose findings s/he is reporting each time, and that this was done recently and successfully, results in the undisputed conclusion that the researcher knows and remembers well what s/he is talking about. Given that, an SI interpretation of the researcher's statements would be less likely to arise. In this new setting, participants were asked to rate how well the interviewer has understood the researcher's statement. They did so on a -3 to +3 Likert scale, where -3 is 'the claim is not understood' and +3 is 'the claim is understood'.⁸ Lastly, the instructions also included four practice items, two instructed to be rated making use of the left part of the scale (semantic contradictions) and two towards the right part (semantically and pragmatically well-formed items).

The vast majority of the items, and specifically the researcher's statements, along with the practice items used in this experiment were adapted from the afore-mentioned work by J. Dotlačil and R. Nouwen. Their items were already constructed so as to fit the researcher-interviewer context, but the interviewer's questions were modified in this experiment so as to serve its own goal. As a result, researcher's statements and interviewer's questions in Experiment 2 were constructed in the exact same way as A's statements and B's questions in Experiment 1, respectively (see indicatively an example of an experimental item from Experiment 2 in (17)).

⁸In Experiment 1 participants turned out not to make use of all points of the scale, so it was considered that that long a scale is not needed and that a shorter scale would work as well. This is the reason why I employed a shorter one this time. Note that the exact same scale is effectively used in the ongoing work by J. Dotlačil and R. Nouwen (p.c.).

- (17) **Onderzoeker:** *Tijdens het evenement werd elke straat door minstens zes agenten beveiligd.*
 researcher during the event was every street by at.least six policemen guarded
 ‘Researcher: During the event every street was guarded by at least six policemen.’
Interviewer: *Werden ze allemaal door evenveel agenten beveiligd?*
 interviewer were they all by as.many policemen guarded
 ‘Interviewer: Were they all guarded by the same number of policemen?’

Moreover, as in Experiment 1, semantically contradictory and pragmatically ill-formed/ infelicitous control items were included, as well as semantically and pragmatically well-formed control items. Crucially, one more condition was added, in which the numeral modifier that interacts with the universal nominal quantifier was *n of meer* ‘*n* or more’. N OR MORE items only differed from AT LEAST (or MORE THAN) items in the type of numeral modifier. Let us now see what this addition serves for. *N or more* and *at least n* express the same numerical relation, both involving non-strict comparison (see Nouwen 2008), and they at the same time exhibit the same behaviour, that is, they trigger the same inferences (namely, SI inferences), and those inferences are weakened when the numeral modifiers in question appear in certain embedded environments. For this reason, Nouwen (2010) puts them in the same class of numeral modifiers, i.e., the so-called class B. As is obvious, *n or more* comes in a disjunctive form, so being similar to *at least n*, it perfectly spells out Coppock and Brochhagen’s (2013) and Büring’s (2008) alternative-based/disjunctive analysis for *at least*. Thus, by adding this condition I aim to test whether those two numeral modifiers will perform similarly, as expected according to the pragmatic view mainly discussed in this paper.

There were two factors in Experiment 2: (i) as in Experiment 1, CONDITION was one factor with N OR MORE as an additional level, and (ii) the type of question the interviewer asked (with two levels). The latter factor is not relevant for the present study, so I will not be discussing it here; the same holds for the MORE THAN condition. The task had six experimental items (three conditions: AT LEAST, N OR MORE, and MORE THAN) and six control items with disjunction (two conditions: semantic contradictions and pragmatic infelicities). Each type of items was rotated through six lists, so that each participant saw one item per condition. Moreover, 13 semantically and pragmatically well-formed control items were included in every list as well as six fillers (similar to those in Experiment 1). Every list had 29 stimuli. 97 filled questionnaires were collected; 18 were excluded, because less than half of each one was filled in, another five were excluded because those subjects were not native speakers of Dutch, and finally, data of six subjects were left out too due to mistakes in the practice items. The final number of observations was $N=1564$.⁹ Last, subjects filled in the questionnaire voluntarily and were naive with respect to the purpose of the study.

⁹ Part of the data of a semantically contradictory control item and of a semantically and pragmatically well-formed control item was not included in the subsequent analyses because of a typo. However, they were not discarded altogether, because the typo was noticed and corrected in time, and the relevant link to the respective list of the experiment was made available again.

3.3.2. Results

As in Experiment 1, SCORES were ordered categorical, thus the data were analysed with ordered probit models. AT LEAST was the reference level of CONDITION and was compared to N OR MORE, MORE THAN and all three types of controls. Intercept random effects for subjects and items were also included in the model.

AT LEAST was rated significantly lower than the well-formed control items ($\beta = -2.168, SE = .214, p < .0001$) and significantly higher than the semantic contradictions ($\beta = .628, SE = .273, p = .022$). As is also evident from Figure 2 on the next page, the difference between AT LEAST and the pragmatic infelicities has vanished ($\beta = .082, SE = .264, p = .756$). This result suggests that the modification of the instructions so as to hinder the interference of the SI readings had an effect. Lastly, no significant difference was found between AT LEAST and N OR MORE ($\beta = .0159, SE = .191, p = .404$). Turning now to the latter condition, quite similarly to AT LEAST's performance, N OR MORE got significantly lower scores than the well-formed control items ($\beta = -2.368, SE = .229, p < .0001$), and its difference from the semantic contradictions (see relevant boxes in Figure 2) was marginally significant ($\beta = .482, SE = .285, p = .091$), while it was not found to be significantly different from the pragmatic infelicities ($\beta = -.07, SE = .277, p = .801$). Further analyses revealed that the difference between semantic contradictions and pragmatic infelicities was significant ($\beta = -.554, SE = .207, p = .008$), with the latter receiving in general higher ratings than the former (see Figure 2), and so was the difference of both the semantically contradictory and the pragmatically ill-formed/infelicitous control items from the well-formed control items ($\beta = -2.96, SE = .262, p < .0001$, and $\beta = -2.344, SE = .242, p < .0001$, respectively), which scored at the upper part of the scale (see Figure 2), as expected and similarly to Experiment 1.

3.3.3. Discussion

The highly significant difference attested between AT LEAST and the well-formed control items, depicted in Figure 2, shows that subjects did not consider the interviewer's questions completely reasonable reactions to the researcher's statements. Having already discussed that this would happen in the case that the researcher's (or A's) statements have in fact given rise to variation effects, we safely draw the conclusion that variation effects do exist with *at least*, replicating the relevant result of Experiment 1, and thus, confirming the first prediction according to the pragmatic view. The second significant finding for *at least*, concerning the difference between the AT LEAST box and that of semantic contradictions one sees on Figure 2, is interpreted as follows: subjects chose not to judge the AT LEAST items similarly to contradictions, but they rather found the former significantly better in reasonableness' terms than the latter. From that I conclude that the unreasonableness or inconsistency caused due to the generation of variation effects in the AT LEAST items is less strong than the one in the case of semantic contradictions, which suggests that those effects are

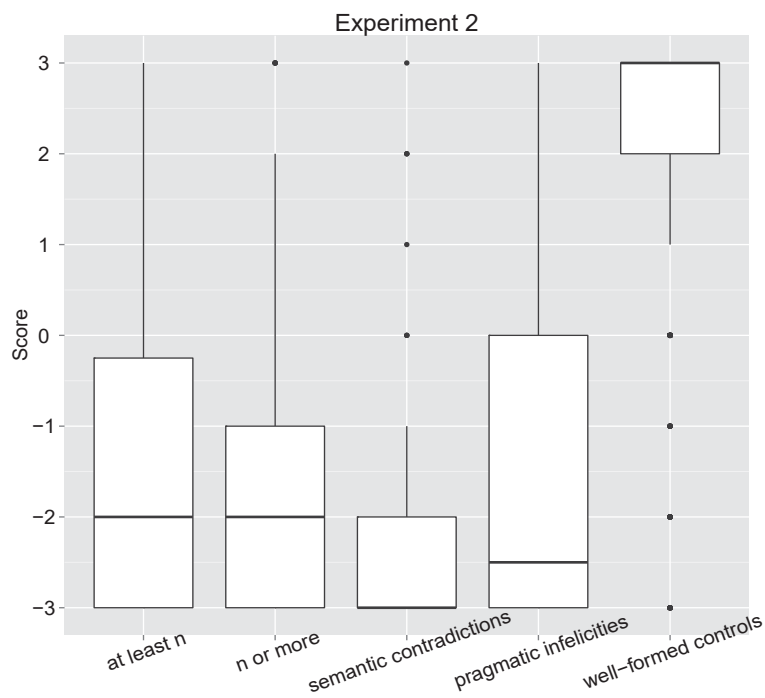


Figure 2: Boxplots of scores per condition. Experiment 2.

pragmatic inferences. This comes to confirm the second prediction of the pragmatic view and also to reinforce the similar finding in Experiment 1. To sum up the findings so far, variation effects arise when *at least* interacts with a universal nominal quantifier and they seem to be generated via a pragmatic mechanism.

As far as *N OR MORE* is concerned, we observe a quite similar behaviour to *AT LEAST*, as predicted. Thus, considering the differences found between *N OR MORE* and the semantically and pragmatically well-formed controls on the one hand, and its marginal difference from the semantic contradictions on the other hand, following the same rationale as above, I conclude that variation effects are available with *n or more* and there are indications that they are not semantically encoded, but rather arise as pragmatic inferences when in the scope of a universal nominal quantifier. It is worth noting that the fact that the comparison of *N OR MORE* with the semantic contradictions approached, although not reached, significance could perhaps be due to an effect of the low scores (contradictory and infelicitous) control items with disjunction obtained. In other words, subjects considering *N OR MORE* items as disjunction-seeming, treated them with lower scores, because they gave that kind of scores to the other inconsistent disjunctive items too.

Furthermore, the absence of a significant difference between *N OR MORE* and *AT LEAST* or the pragmatically ill-formed control items with disjunction is compatible with an alternative-introducing semantics for *at least* à la Büring (2008) and Coppock and Brochhagen (2013). Moreover, it is in

line with *at least* and *n or more* sharing a similar mechanism generating variation effects; similar to the fc implicature-generating mechanism with disjunction (cf. Alonso-Ovalle 2005).¹⁰

4. Conclusion

In this paper I investigated experimentally the availability of variation inferences, akin to free choice inferences, which are triggered when lower-bound class B numeral modifiers, such as *at least* and *n or more*, appear in the scope of a universal nominal quantifier. In addition, I tested the semantic/pragmatic strength of those effects, that is, whether they arise as semantic or as pragmatic inferences. According to the dominant view on those inferences, i.e., the pragmatic view, they arise in certain embedded contexts and, in fact, they do so via an implicature-generating mechanism. This study, by means of two experiments, such that one replicated the results of the other, contributes evidence for the pragmatic strength of variation effects with *at least*, thereby providing strong support in favour of the pragmatic view on the derivation of those effects. Further findings this study makes available are in line with an alternative-introducing semantics for *at least* in the style of Büring (2008) or Coppock and Brochhagen (2013), and are also compatible with a common pragmatic mechanism responsible for the derivation of variation inferences with *at least* and of free choice implicatures with disjunction.

References

- Alonso-Ovalle, L. (2005). Distributing the disjuncts over the modal space. In Bateman, L. and C. Ussery (Eds.), *Proceedings of the 35th North East Linguistics Society Conference*, Volume 35, pp. 75–86. University of Massachusetts, Amherst: GLSA.
- Alonso-Ovalle, L. and P. Menendez-Benito (2010). Modal indefinites. *Natural Language Semantics* 18, 1–31.
- Büring, D. (2008). The least *at least* can do. In Chang, C. B. and H. Haynie (Eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*. Cascadilla Proceedings Project.
- Christensen, R. (2013). Ordinal—regression models for ordinal data. R package version 2013.9-30.
- Ciardelli, I., J. Groenendijk, and F. Roelofsen (2012). Inquisitive semantics. NASSLLI lecture notes.
- Coppock, E. and T. Brochhagen (2013). Raising and resolving issues with scalar modifiers. *Semantics & Pragmatics* 6, 1–57.
- Cummins, C. and N. Katsos (2010). Comparative and superlative quantifiers: Pragmatic effects of comparison type. *Journal of Semantics* 27, 271–305.

¹⁰See also Kratzer and Shimoyama (2002), considering their example with *irgendein* in (8), with a disjunctive relevant set of alternatives.

- Doran, R., R. Baker, Y. McNabb, M. Larson, and G. Ward (2009). On the non-unified nature of scalar implicature: An empirical investigation. *International Review of Pragmatics 1*, 1–38.
- Geurts, B. and R. Nouwen (2007). At least et al.: the semantics of scalar modifiers. *Language 83*, 533–559.
- Kamp, H. (1973). Free choice permission. In *Proceedings of the Aristotelian Society*, Volume 74, London, pp. 57–74. Methuen.
- Klinedinst, N. (2007). *Plurality and Possibility*. Ph. D. thesis, UCLA.
- Kratzer, A. and J. Shimoyama (2002). Indeterminate pronouns: The view from Japanese. In Otsu, Y. (Ed.), *Proceedings of the 3rd Tokyo conference on psycholinguistics*, pp. 1–25.
- Krifka, M. (2009). Approximate interpretations of number words: A case for strategic communication. In Hinrichs, E. and J. Nerbonne (Eds.), *Theory and Evidence in Semantics*, pp. 109–132. Stanford: CSLI Publications.
- Mayr, C. (2013). Implicatures of modified numerals. In Caponigro, I. and C. Cecchetto (Eds.), *From grammar to meaning: the spontaneous logicity of language*, pp. 139–171. Cambridge: Cambridge University Press.
- McNabb, Y. and D. Penka (2014). The interpretation of Superlative Modifiers and Deontic Modals: An Experimental Investigation. In Ettxeberria, U., A. Fălăuș, A. Irurtzun, and B. Leferman (Eds.), *Proceedings of SuB 18*, pp. 271–288.
- Nickel, B. (2010). Generically free choice. *Linguistics and Philosophy 33*, 479–512.
- Nouwen, R. (2008). Upperbounded no more: the exhaustive interpretation of non-strict comparison. *Natural Language Semantics 16*, 271–295.
- Nouwen, R. (2010). Two kinds of modified numerals. *Semantics and Pragmatics 3*, 1–41.
- Nouwen, R. (2015). Modified Numerals: the Epistemic Effect. In Alonso-Ovalle, L. and P. Menendez-Benito (Eds.), *Epistemic Indefinites*, pp. 244–266. Oxford University Press.
- Schwarz, B. (2013). At least and Quantity Implicature: Choices and Consequences. In Aloni, M., M. Franke, and F. Roelofsen (Eds.), *Proceedings of the 19th Amsterdam Colloquium*, pp. 187–194.
- van Tiel, B., E. van Miltenburg, N. Zevakhina, and B. Geurts (2014). Scalar diversity. *Journal of semantics* (doi:10.1093/jos/ffu017).
- Zimmermann, T. E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics 8*, 255–290.