

Differences and similarities between scalar inferences and scalar modifiers: The case of quantifiers *

Yaron McNabb
Utrecht University

Abstract Despite the rich theoretical and experimental work on scalar implicature, many of the studies on this topic were limited to *some* vs. *all*, neglecting the cross-categorical pervasiveness of the phenomena. The few experimental studies involving a more diverse group of scalar implicatures have found variation among expressions in the likelihood they give rise to scalar implicature, thereby challenging the assumption that scalar implicature (and generalized conversational implicature) is a uniform phenomenon (Doran, Baker, McNabb, Larson & Ward 2009; Doran, Ward, Larson, McNabb & Baker 2012; van Tiel, van Miltenburg, Zevakhina & Geurts 2014). This paper presents a first, systematic investigation of the degree to which a large group of quantifiers give rise to the implicature ‘not all’ using an utterance compatibility task with a modified Likert scale. Two accounts for the variation among quantifiers are proposed: (i) Shared semantic properties among three coherent groups of quantifiers account for the degree they give rise to upper-bound implicature; or (ii) the likelihood of an implicature is a function of the scalar distance between the various quantifiers and ‘all.’ The predictions these two accounts make are discussed, charting the way to a future investigation of the heterogeneity of scalar implicature.

Keywords: scalar implicature, quantifiers, superlative modifiers, scalar diversity, experimental pragmatics

1 Introduction

The scalar expression *some* in (1) is assumed to evoke a set of alternatives whose members are ordered in terms of informativeness along the lines of (2). The use of *some* implicates that informationally-stronger alternatives like *most* or *all* don’t hold (Horn 1972; Gazdar 1979; Levinson 2000; Geurts 2010).

* This paper has greatly benefited from discussions with Stavroula Alexandropoulou, Dominique Blok, Judith Degen, Jakub Dotlačil, Nick Fleischer, Dan Lassiter, Roger Levy, and above all Rick Nouwen, as well as audiences at SALT 25, XPRAG 2015 and the ELiTU series at Utrecht University. 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.

- (1) **Some** of Ann's friends shared her view of the world as a dark place.
 (2) <all, most, many, some>

Despite the rich theoretical and experimental work on scalar implicature, many of the studies on this topic were limited to *some* vs. *all*, neglecting the cross-categorical pervasiveness of the phenomena (but see Horn 1972 and Hirschberg 1985). The few experimental studies involving a more diverse group of scalar implicatures have found variation among expressions in the likelihood they give rise to scalar implicature, thereby challenging the assumption that scalar implicature (and generalized conversational implicature) is a uniform phenomenon (Doran et al. 2009, 2012; van Tiel et al. 2014).

The main contribution of recent studies on the diversity of scalar implicature to theoretical and experimental semantics and pragmatics is the finding that when looking at a varied group of scalar expressions, additional linguistic and extralinguistic factors that increase the likelihood of an implicature emerge. This finding provides a foundation to further explore these and other factors on heretofore neglected expressions. Consider, for example, quantifiers: Doran et al. (2009) predict that these expressions would be less susceptible to contextual factors such as salience of alternates on a scale than, say, gradable adjectives. van Tiel et al. (2014) show that the reliable factors are rather whether the strongest alternate, e.g., *all*, is a maximal value, and the extent to which the quantifier in the utterance, e.g., *a handful* or *most*, is closely associated with, and more distant (on the scale) than, the stronger alternate *all*. The scope of both of these studies, however, meant that while a large and diverse group of scalar expressions were examined, only a small number of members within each linguistic category, e.g., quantified NPs, were tested, namely only *some*, *most* and *all*.

In this study, 11 different quantifiers were examined in order to determine whether what Doran et al. (2009) and van Tiel et al. (2014) found for a couple quantified NPs is generalizable to a larger sample of these expressions. The rate of scalar implicature the different quantifiers give rise to was tested in an utterance compatibility judgement task, where participants evaluated a CLAIM with the quantifier vis-à-vis a FACT with *all*. The results point to differences among quantifiers with respect to the extent to which they give rise to a scalar implicature. Fractions like *quarter* are the most likely to give rise to scalar implicature, whereas existential quantifiers like *a handful* are less likely, and vague quantity expressions like *a great deal* are the least likely to give rise to an implicature.

2 Continuous vs. binary data and semantic vs. pragmatic upper bounds

Many studies in experimental semantics and pragmatics use a binary truth value judgement task. An alternative to the binary task is a one that employs a Likert scale or a continuous truth/false scale (or a version thereof). Each of these methodological choices play a role in shaping participants' responses. In the binary task, for example, participants are forced to make sharp judgements even when their responses may be more nuanced. Moreover, the binary understanding of truth and falsity in semantic theories is a theoretical concept that may not correspond to speakers' understanding of 'truth' and 'false' (Hansen & Chemla 2013).

Responses from tasks using the Likert or the continuous truth/false scale are meaning-measuring methods that should likewise be approached with great care. We don't know whether speakers start off with a gradient notion of truth and falsity, and if they do, we don't know what factors tilt a proposition or an utterance toward the true or false end of such a scale. What we can do, as many researchers have, is use a scale in order to allow for a gradient response and examine whether significant contrasts in responses to different experimental conditions shed light on different types of meaning, in this case semantic vs. pragmatic meaning. It's crucial to point out that no point on the scale should be interpreted as mapped onto a specific type of meaning (e.g., logical or pragmatic meaning). That said, control items that are unequivocally false or blandly true are expected to cluster around opposite end-points on the scale.

In this study, I adapt the part of Doran et al.'s task which asks participants to contrast a CLAIM and a FACT. Instead of asking participants whether the CLAIM was true, given the FACT, I asked participants to rate how compatible the CLAIM was with the FACT. Participants were asked to rate compatibility on a scale from -3 to 3 , where -3 was completely incompatible and 3 was completely compatible. Such scales have been successfully used in Cummins & Katsos 2010 and McNabb & Penka 2014, *et seq.* The intuition behind using a scale from a negative to a positive number is that participants have shown to map three different types of semantic-pragmatic status of an utterance on three different regions on the scale: semantically- and pragmatically-felicitous utterances were mapped around the positive end of the scale and semantically-anomalous or logically false utterances were mapped on the negative end of the scale, both types of utterances showing a small variance. The third type of utterances, utterances that were semantically well-formed but pragmatically infelicitous, usually exhibited a large variance. It should be reiterated that the exact points on the scale don't (always) matter; rather, contrasts in the response patterns is what I assume should be a sign of differences between different types of meaning.

The extent to which participants draw a scalar implicature is usually experi-

mentally determined vis-à-vis baseline or control items for which responses are predicted to be at floor or ceiling with respect to the task. But very few studies make a direct comparison between an entailed upper bound and an implicated one. Modification by *at most* is a reliable way to impose an entailed—as opposed to an implicated—upper bound. Krifka (1999) and Geurts & Nouwen (2007) observe that *at most* imposes an uncancellable upper-bound on numeral, scalar and non-scalar expressions. This is illustrated in (3-5), in which the upper-bound can be suspended in the non-modified (a) sentences but not in the modified (b) sentences.

- (3) a. Last night, Willow tested **three** new spells (, if not **four**).
 b. Last night, Willow tested **at most three** new spells (#, if not **four**).
- (4) a. Buffy read **a handful** of the books on the shelf (, if not **all** of them).
 b. Buffy read **at most a handful** of the books on the shelf (#, if not **all** of them).
- (5) a. Dr. Walsh is **an assistant professor** (, if not **an associated professor**).
 b. Dr. Walsh is **at most an assistant professor** (#, if not **an associated professor**).

In the experiment reported in Section 3, I used quantified NPs modified by *at most* as a baseline and compared it to the likelihood of upper-bound construal in the non-modified cases. On the assumption that modification by *at most* leads to a semantic upper bound whereby the modified term is interpreted as the maximal value on a scale, significantly different and varied rates for the non-modified quantifier in comparison can be interpreted as a cancellable (i.e., pragmatic) upper bound.

3 Methods

3.1 Preliminaries

Instead of constructing stimuli or drawing upon examples from the literature, I gleaned 11 naturally-occurring utterances from the Corpus of Contemporary American English (COCA, Davies 2008), looking for quantifiers modified by *at most*. The 11 quantifiers, given in (6), were not chosen for their frequency.

- (6) The 11 quantifiers used in the experiment:
- | | | | |
|-------------|-----------------|--------------|---------|
| a. a bunch | d. a great deal | g. a number | j. half |
| b. a couple | e. a handful | h. a quarter | k. most |
| c. a few | f. a multitude | i. a third | |

For each of the 11 utterances, two factors were manipulated: (i) modification by *at most* (or not) and (ii) the quantifier used (Q in (7)).

(7) 1. Example utterance:

{Q / At most Q} of Ann's friends shared her view of the world as a dark place.

In order to ensure all 11 utterances were well-formed with all 11 quantifiers, both modified and not modified, all the utterances were rated for well-formedness by 10 participants on Amazon's Mechanical Turk. All scored between 4 and 6 on a 7-point Likert scale ranging from 0 to 6 (mean=5.04; SD=.81; median=5).

3.2 The task and procedure

Participants on Mechanical Turk rated the compatibility of pairs of CLAIMs and FACTs on a modified Likert scale ranging from -3 to 3. In addition to the two modification conditions (non-modification and *at most*) and the 11 quantifier conditions, there were also three discrepancy conditions that differed in the term used in the follow-up, FACT statement: *none*, *some* and *all*, as illustrated in (8).

(8) Example stimulus:

Claim: {A / At most a} a handful of Ann's friends shared her view of the world as a dark place.

Fact: {None / Some / All} of Ann's friends shared her view of the world as a dark place.

How compatible is the CLAIM with the FACT?

-3	-2	-1	0	1	2	3
completely incompatible						completely compatible

I created 66 lists using a Latin Square design so that no participant saw the same kind of combination of modification, quantifier and discrepancy condition for a given sentence (=11 sentences × 11 quantifiers × 2 modification conditions × 3 discrepancy conditions). Each of the 66 lists were evaluated by two different participants (total of N=132). This allowed for 6 observations for each quantifier × modification × discrepancy per person.

4 Predictions

Different predictions can be made about the compatibility between the CLAIM and the three discrepancy conditions in the FACT depending on whether the quantifier in

question is modified by *at most* or not. Let's consider the quantifier *a handful* as an example. Horn's (2001: 235–236) “suspender test” in (9–10) shows that *a handful* asymmetrically entails *some*, given the contrast between the well-formed sentences in (9a) and (10a) and the ill-formed sentences in (9b) and (10b). Given that *a handful* entails *some*, it's incompatible with *none*.

- (9) a. I care about you Naruto. And I'm sure there's **some if not a handful** of people who do too.¹
 b. # I care about you Naruto. And I'm sure there's **a handful of, if not some**, people who do too.
- (10) a. Many of them are hardheaded and ‘makulit’ but **some, if not a handful** of them are amenable to change.²
 b. # Many of them are hardheaded and ‘makulit’ but **a handful, if not some** of them are amenable to change.

The case is less clear for *at most a handful*, which creates a downward-monotone environment, licensing NPIs, as shown in (11). If the entailment pattern of *at most a handful* is similar to other downward-monotone quantifiers like *few*, then *at most a handful* is compatible with *none*. Moreover, we can suspend the purported existential presupposition, as in (12), if we introduce speaker ignorance into the utterance. Krifka (1999), however, notes that *at most Q* can introduce discourse referents—albeit not as successfully as *at least Q*—suggesting that there may be some cases in which *at most* carries what looks like an existential presupposition, as illustrated in (13). I return to this issue in Section 6.3.

- (11) I have read or viewed hundreds of news stories about the Pain-Capable Unborn Child Protection Act over the past six months, and **at most a handful, or less, have made any mention** of the Gonzales ruling[. . .]³
- (12) From what I remember, **at most a handful** of the news stories, if any, have made any mention of the Gonzales ruling.
- (13) a. ?? **At most three** boys left. They found the play boring.
 b. **At least three** boys left. They found the play boring.

Finally, the critical conditions are the ones where the FACT contains *all*. While *a handful* may be used to implicate ‘not all,’ *at most a handful* entails it. Table 1 summarizes the predictions for each experimental condition.

1 Source: <http://www.bestfanfictions.com/s/3403741/2/Demon-s-Revenge>

2 Source: <http://www.rcmanila.org/sites/rcmanila.org/files/balita/MARCH-27-2014-BALITA.pdf>

3 Source: <http://www.lifenews.com/2015/06/11/american-college-of-obstetricians-and-gynecologists-has-become-a-de-facto-pro-abortion-group/>

CLAIM	FACT	Compatible?	Reason
A handful ...	None ...	No	Entails <i>some</i>
At most a handful ...	None ...	?	DE/Existential presupposition
A handful ...	Some ...	Yes	Entailment relation
At most a handful ...	Some ...	Yes	Entailment relation
A handful ...	All ...	No	Implicature
At most a handful ...	All ...	No	Contradiction

Table 1 CLAIM-FACT compatibility in the various experimental conditions

5 Results

The results of the experiment were analyzed with a mixed-effects ordered probit regression model with random effects for subjects and items and fixed effects for modifier, discrepancy and quantifier, using the Ordinal package (Christensen 2012) on R. As can be seen in Figure 1, overall scores for claims with the quantifiers modified by *at most* were significantly lower than scores for those with non-modified quantifiers ($\beta=1.36$, $SE=0.14$, $p<0.001$). Scores for the *all* discrepancy condition were significantly lower than those for the *none* discrepancy condition ($\beta=0.35$, $SE=0.15$, $p<0.05$) and the *some* discrepancy condition ($\beta=3.45$, $SE=0.17$, $p<0.001$). CLAIMs with *none* following FACTs with a quantifier modified by *at most* were rated higher than when following FACTs with a non-modified quantifier ($\beta=-4.32$, $SE=0.26$, $p<0.001$) and CLAIMs with *some* following FACTs with a non-modified quantifier ($\beta=-0.61$, $SE=0.20$, $p<0.01$).

5.1 Quantifiers modified by *at most*

Zooming in on CLAIMs with quantifiers modified by *at most*, FACTs with *all* received significantly lower scores than FACTs with *none* ($\beta=0.42$, $SE=0.17$, $p<0.001$) and FACTs with *some* ($\beta=4.10$, $SE=0.15$, $p<0.001$). Most of the results are as expected, as *all* is incompatible with any of the 11 quantifiers when modified by *at most*. *Some*, on the other hand, is expected to be compatible with all these quantifiers when they're modified by *at most*. Note that although the scores for *none* were low, they were reliably different than those for *all*. This distinction will be discussed in more detail in Section 6.3. There was no main effect of quantifier and no differences between individuals quantifiers ($p>0.05$ for all comparisons).

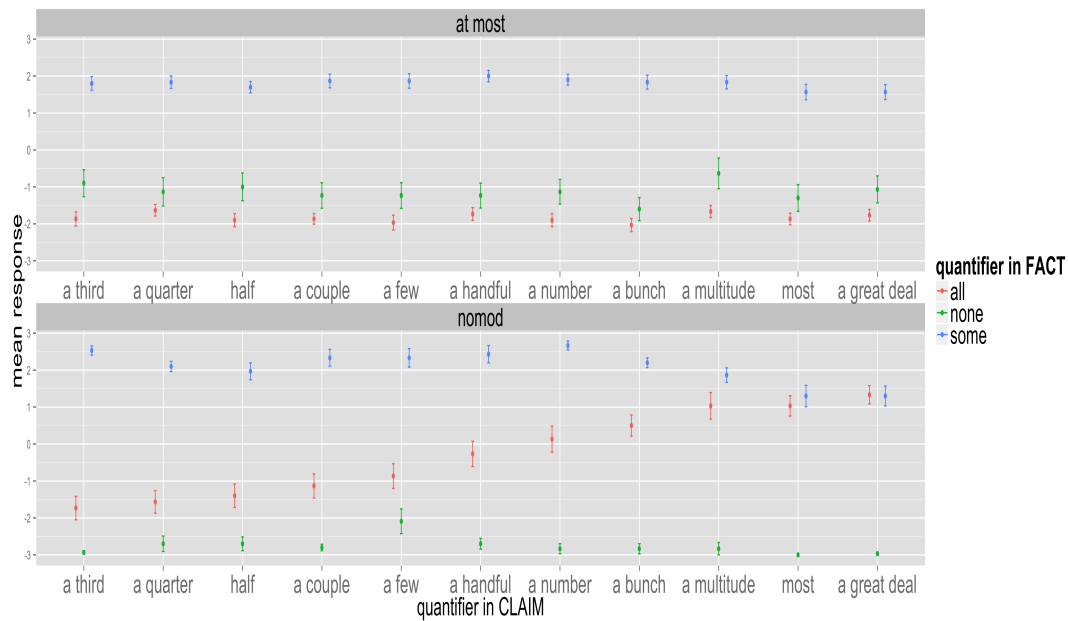


Figure 1 Mean responses (and standard errors) of compatibility of CLAIMs with FACTs, per quantifier, per modification condition.

5.2 Non-modified quantifiers

For CLAIMs with non-modified quantifiers, FACTs with *none* received significantly lower rates than FACT with *all* ($\beta=3.54$, $SE=0.1$, $p<0.001$) and FACTs with *some* ($\beta=6.12$, $SE=0.15$, $p<0.001$). As expected, *none* was incompatible with all of the 11 quantifiers and *some* was compatible with all of them.

The scores for *all* followed by the non-modified quantifiers exhibit a great deal of diversity. Pairwise comparisons point to a clustering of three groups: The first group, including *a third* (mean=-1.73), *a quarter* (mean=-1.57), *half* (mean=-1.40), and *a couple* (mean=-1.13), received the lowest compatibility ratings with *all*. The second group, including *a handful* (mean=-.27), *a number* (mean=-.13) and *a bunch* (mean=.50), received middling compatibility ratings with *all*. And the third group, including *most* (mean=1.03), *a multitude* (mean=1.03) and *a great deal* (mean=1.33), received the highest compatibility ratings with *all*. The quantifier *a few* (mean=-.87) wasn't significantly different than all of the quantifiers in the first or the second group and so falls between the two with respect to its compatibility with *all*.

6 Discussion

The results of the experimental study of a heterogeneous group of quantifiers provide evidence for the diversity with which participants contrast these quantifiers with *all* as a way to gauge the extent to which each one of these quantifiers leads to an upper-bound construal. While Doran et al. (2009) and van Tiel et al. (2014) already show that various scalar expressions, e.g., quantifiers, adjectives and modals, exhibit different likelihoods of implicature generation, this study, by focusing on quantifiers only, shows that even within a smaller group of scalar expressions, the extent to which these expressions give rise to an upper-bound is great.

There are two apparent ways to interpret the results reported here: The first possibility is that the three groups of quantifiers that emerged—fractions, existential quantifiers, and vague quantifiers of large degree—differ in their likelihoods to give rise to scalar implicature due to their contrastive lexical semantics, as shown in (14). The second possibility is that the division into groups is a misinterpretation of the data and in fact the extent to which a scalar implicature is generated is a function of the scalar distance of the quantifier from the upper-bound of the scale associated with it, i.e. *all*, as shown in (15).

- (14) Three groups w.r.t. to their compatibility with *all* from least to most compatible:

{a third / a quarter / half / a couple} \prec^4 {a handful / a number / a bunch} \prec
 {a multitude / most / a great deal}

- (15) Compatibility with *all* from least to most - merely ordered:

a third \prec a quarter \prec half \prec a couple \prec a few \prec a handful \prec a number \prec a bunch \prec a multitude \prec most \prec a great deal

6.1 Lexical semantics feeds pragmatic upper-bound construals

Recall (and consult (14)) that the quantifiers that were the least compatible with *all* were *a third*, *a quarter*, *half* and *a couple*, followed by *a handful*, *a number* and *a bunch* (with *a few* straddling the two groups), and that the quantifiers that were the most compatible with *all* were *a multitude*, *most* and *a great deal*. Do the quantifiers in each group that emerges from the results share any other properties other than likelihood of scalar implicature? If so, can these properties be attributed to the different likelihood of scalar implicature each group of quantifiers exhibits? In the next sections I show that these three groups of quantifiers in fact share a few semantic characteristics that may suggest that they are semantically-coherent groups.

⁴ The quantifier *a few* resides somewhere in the statistical twilight zone between the first and second group.

6.1.1 Fractions

Note that the first group included mostly fractions: *a third*, *a quarter*, *half* with the exception of *a couple*, which is ambiguous between as ‘two’ and ‘small quantity approximating two.’ The central shared property of these quantifiers is that they are all ambiguous between a precise and approximate reading. In order to resolve this ambiguity, they can be modified by the slack regulator *exactly* or *approximately*, in contrast with the quantifiers in the other two groups (Lasersohn 1999; Sauerland & Stateva 2007).⁵ What this suggests is that fractions and *a couple* denote an upper-bound that can either be mapped onto a range (the approximate reading) or an exact fraction or number (the exact reading). The levels of granularity determines whether we get the approximate or the exact reading, but in either reading there is an upper-bound that excludes *all* (and larger fractions on the scale).

- (16) a. {exactly / approximately} {a third / a quarter / half / a couple}
 b. # {exactly / approximately} $\left\{ \begin{array}{ll} \text{a bunch} & \text{a few} \\ \text{a great deal} & \text{a handful} \\ \text{a multitude} & \text{most} \\ \text{a number} & \end{array} \right\}$

The move to analyze fractions as having an upper-bound is similar to the growing evidence and arguments for a two-sided analysis of numbers in contrast with the classic view, in which number words have a one-sided semantics (Horn 1992; Musolino 2004; Geurts 2006; Huang, Spelke & Snedeker 2013; Kennedy 2013). One of the tests in which numbers and fractions pattern together to the exclusion of *some* is the negation test give below (Kennedy 2013). On the assumption that the negation in the first clause can either target the semantic content or the pragmatically-enriched utterance, but not both, saying in (17) that neither Buffy nor Willow read many books cannot be followed by saying that Buffy read some, negating the semantic content, and that Willow read all, negating the pragmatically-enriched content. Such a continuation with numbers words, as in (18), is felicitous, strongly suggesting that the two-sided meaning of numbers is semantic and so can be negated and corrected without the zeugma effect in (17). Fractions seem to pattern with number words: The coherent continuation in (19) suggests that the upper-bound reading of *third* is part of the semantic content, and so *a quarter* and *all* are both likely corrections.

- (17) # Neither Buffy nor Willow read **many** of the books; Buffy read **some** and Willow read **all**.

⁵ Some of the quantifiers in the other two groups are marginally felicitous when modified by *approximately*, possibly because this modifier reinforces the approximate reading. None of them, however, is compatible with *exactly*.

- (18) Neither Buffy nor Willow fought **three** vampires; Buffy fought **two** and Willow fought **four**.
- (19) Neither Buffy nor Willow lost **a third** of their respective weapon arsenals; Buffy lost **a quarter** of it and Willow lost **all** of it.

6.1.2 Existential quantifiers and their ilk

The second group of quantifiers with respect to the likelihood of implicature generation are the existential quantifiers *a handful*, *a number* and *a bunch*. These quantifiers were significantly more compatible with *all* than the fractions discussed in Section 6.1.1. Unlike these three quantifiers, the quantifier *a few* wasn't significantly different than all fractions or all the vague quantity expressions (e.g., *a multitude*) but will be discussed here together with the other existential quantifiers.

How do these existential quantifiers differ from the existential quantifier *some*? As already shown in Section 4, Horn's (2001) suspender test, illustrated in (9-10), suggests that *a handful* asymmetrically entails *some*. Applying this test to the other existential quantifiers in (20-22) shows that they all asymmetrically entail *some*, too.⁶

- (20) a. When we stayed here there were **some if not a few** children there as expected, but none of them were running wild as the other comments state.⁷
b. # When we stayed here there were **a few if not some of the** children there as expected, but none of them were running wild as the other comments state.
- (21) a. Despite my wish to use **some if not a number** of these suggestions combined, my business partner vetoed my desire.⁸
b. # Despite my wish to use **a number if not some** of these suggestions combined, my business partner vetoed my desire.
- (22) a. I'm really hoping that **some, if not a bunch** of us, can get together and have some fun that weekend and hang out and have a little get together!⁹
b. # I'm really hoping that **a bunch, if not some** of us, can get together and have some fun that weekend and hang out and have a little get together!

These existential quantifiers thus share the basic existential meaning of *some* and encode in addition something like a 'small quantity' meaning. I leave the question of

⁶ The same kind of contrast is observed when *but not* or *or even* is substituted for *if not*.

⁷ Source: http://www.tripadvisor.co.uk/ShowUserReviews-g662606-d291481-r135574713-Costa_Adeje_Gran_Hotel-Costa_Adeje_Adeje_Tenerife_Canary_Islands.html

⁸ Source: <http://steelwhitetable.org/2011/06/24/contest-name-new-popcorn-business>

⁹ Source: <https://www.facebook.com/theCHIVEReno/posts/454623144689389>

whether the small quantity meaning can be captured with a semantics similar to the one Solt (2009) proposes for *few*, but if we use the suspender test again, contrasting these quantifiers with *many*, which denotes a contextually-determined large quantity (Solt 2009), we see that they implicate ‘not many.’

- (23) a. # Many people, if not/but not { a number / a bunch / a handful / a few }
 b. { A number of / A bunch of / A handful of / A few } students came if not/but not many.

Quite a few experimental studies show that while the implicature ‘some but not all’ for utterances with *some* is frequent, the scalar implicature isn’t always generated (Bott & Noveck 2004; Guasti, Chierchia, Crain, Foppolo, Gualmini & Meroni 2005; Pouscoulous, Noveck, Politzer & Bastide 2007, *inter alia*). This is similar to the middling compatibility ratings of statements with these quantifiers with FACTs with *all* (means of –1 to 1).

6.1.3 Vague quantity expressions

The third class of quantifiers with respect to their compatibility with subsequent utterances with *all* includes *a multitude*, *most* and *a great deal*. These quantifiers were the most compatible with FACTs containing *all*. The semantic properties these quantifiers share that may render them more likely to be compatible with *all* are (i) these quantifiers’ denoting a large quantity and (ii) their context sensitivity.

That the quantifiers *a multitude* and *a great deal* denote large quantities is apparent from the requirement that whenever a proposition containing either one is true, the same proposition with *many* substituted for either one must be true, too, as illustrated in (24). The suspender test in (25) and (26) suggests that *a multitude* and *a great deal* entail *many*. It may be the case, then, that these two quantifiers denote something similar to *many* plus an additional emphatic meaning.

- (24) # { **A multitude / A great deal** } of the questions appear to have come from teen library patrons, **but not many** of the questions have come from teen library patrons.
- (25) a. The university has **many, if not a multitude** of ‘lines of flight’ available to it.¹⁰
 b. # The university has **a multitude, if not many** ‘lines of flight’ available to it.
- (26) a. The New York Daily News reports that **many** of the pill bottles, if not **a great deal** of them, were not labeled and were probably older prescriptions.

¹⁰ Source: Barnett, Ronald. 2013. *Imagining the University*. New York, NY: Taylor & Francis.

- b. # The New York Daily News reports that **a great deal** of the pill bottles, if not **many** of them, were not labeled and were probably older prescriptions.

The obvious next question is how *most* fits in this picture. Under the generalized quantifier theory, *Most As are B* is true just in case the cardinality of the As that are B is greater than the cardinality of the As that aren't B, as formalized in (27) (Barwise & Cooper 1981).

$$(27) \quad \llbracket \text{most} \rrbracket (A)(B) = 1 \text{ iff } |A \cap B| > |A - B|$$

However, nothing in this GQT-style denotation says anything about whether the cardinality of $|A \cap B|$ corresponds to a large quantity, and for a good reason. Suppose, for example, that we talk about three types of whiskey that received a prize, and I say that most of these whiskeys were Japanese. You would conclude that two whiskeys out of the three that received the prize were Japanese but wouldn't necessarily conclude that many Japanese whiskeys got a prize, given that there are nine active Japanese distilleries (Baker 2004). And so, even under a decompositional analysis of *most*, the silent counting quantifier (or gradable modifier that modifies plural NPs ranging over pluralities) *many* that combines with the superlative morpheme *-est* (Hackl 2009) is not the quantity adjective *many* (Solt 2009). This point can be further illustrated by the fact that the complement set of *As that are B* can also include a large number above a contextually-salient standard of large quantity, as the continuation with *many* in (28) exemplifies.

- (28) Most students passed the exam, but many didn't.

What the semantics in (27) does achieve, however, is the lower-bound (one-sided) semantics of *most* assumed by many Neo-Griceans (Gazdar (1979) *et seq.*), leaving the upper-bound inference to be derived as scalar implicature. Contrary to the Neo-Gricean approach, Ariel (2004) argues that *most* has a two-sided semantics, specifically "a proper subset which is the largest subset given any partitioning of the complement subset," but argues that this semantics doesn't have an 'all'-exclusion entailment. Ariel's (2004) analysis accounts for the cases in which *most* is interpreted as 'not all,' following the two-sided semantics, as well as the cases in which *most* is compatible with 'all,' arguing that the asserted lexical meaning of *most* is distinct from the truth compatibility of *most* with that of *all*. Under this approach, it's not surprising that responses for *most*, which doesn't implicate 'not all,' are different than responses for, e.g., *a few*, which does implicate 'not all.' The question that remains is that if the notion of compatibility, calculated pragmatically, is distinct from the semantic content, then what drives the variation in compatibility with *all*? If we generalize Ariel's (2004) approach to all quantifiers, then there should be no

difference in their compatibility with *all*, contrary to the results presented in this study.

In conclusion, the group of expressions that were the most compatible with utterances with *all* is a mixed group consisting of two vague quantity expressions and the proportional quantifier *most*. In the case of *a multitude* and *a great deal*, it's possible that since these expressions denote a context-dependent large quantity, *all* isn't a likely alternative in a similar way that *the tallest* isn't a likely alternative to *tall*. Relatedly, Doran et al. (2009) found that scalar implicatures arise less often for gradable adjectives than for other scalar expressions, like quantifiers and modals. *Most*, on the face of it, shouldn't pattern with the vague quantity expressions, given that by virtue of it denoting a proportion of the set, it should evoke an alternative with an expression that takes the entire set, namely *all*. Ariel's (2004) analysis, whereby *most* doesn't exclude *all*, doesn't explain the lower compatibility rates of other quantifiers that don't have an upper bound, e.g., *a few*. It's possible, then, that *all* was interpreted in a hyperbolic way; that is, *all* is used here to mean 'a considerably large quantity,' as in the sentence in (29), uttered by a surprised child that was expecting to eat a large slice of cake but realized that only a thin sliver was left.

(29) You ate all the cake!

6.1.4 Caveats to, and potential pitfalls of, the lexical semantic account

The quantifiers studied here seem to form three groups whose shared characteristics are fairly intuitive: fractions, which denote an upper bound albeit subject to precisification; existential quantifiers, which pattern with *some* with respect to scalar implicature; and vague quantity expressions that denote a large quantity.

But while these three groups seem semantically coherent, a growing number of studies point to lack of uniformity with respect to implicature strength among scalar expressions (Doran et al. 2009; van Tiel et al. 2014), among different types of generalized conversational implicatures (Doran et al. 2012), and even under different structural and discursive contexts for the most studied scalar implicature 'not all' derived from *some* (Degen 2015). And so, providing an account that banks on one factor (semantic properties) is a good start but may be misleading, as the semantic properties sketched in this section may in fact interact with additional, previously-studied or yet uncovered, factors.

6.2 Scalar distance strengthens pragmatic upper-bound construals

van Tiel et al. (2014) found that the reliable factors contributing to the likelihood to draw a scalar implicature are scale boundedness and scale association. Scalar

expressions that lie on a scale whose strongest term is maximal, e.g., <all, some> and <succeed, attempt> are more likely to lead to an implicature than scalar expressions on a scale whose strongest term has a one-sided, lower bound meaning, e.g., <beautiful, good-looking>. ¹¹ In addition, scalar expressions are more likely to lead to the implicature that a stronger term doesn't hold if the two expressions are frequent collocates and if their semantic distance is great.

As the quantifiers studied here can all be argued to lie on an upper-bounded scale with *all* as the maximal alternate, boundedness was controlled for. Estimates of scale association for only some of the quantifiers were found in the Google Web 1T 5-Gram. As given in Table 2, among the quantifiers examined in the experiment, only *most*, *a few* and *half* had high enough frequency ($f \geq 100$) and collocated with *all*, using the structural cues *Q if not/but not/ and perhaps/and maybe all* (see under the N-gram column). The scores for *some* and *many* were added in the table for comparison. It's clear that *most* is the most associated with *all*, with *a few* and *half* trailing behind. Note also that *some* and *many* frequently collocate with *all*.

On the assumption that stronger association between a quantifier and *all* increases the probability to draw the scalar implicature 'not all,' then we would expect *most* to have the highest implicature rate, followed by *some* and *many*, which weren't tested in this experiment, with *a few* and *half* being less likely to give rise to implicatures. The compatibility scores with *all*, however, show the opposite effect: *half* is more likely to give rise to an implicature than *a few* and *most*. That said, a future, more careful calculation of collocation is warranted, ¹² as well as experimental work that explicitly surveys the degree of association as well as perceived distance between the various quantifiers and *all*.

For lack of experimental data on association and distance, I plot in Figure 2 the quantifiers with respect to the lower and upper bound they semantically and pragmatically mean. What this Figure captures isn't the perceived distance in meaning between the various quantifiers and *all* but rather what the range of proportions each one of these quantifiers may denote. I assume that fractions denote the precise meaning with an error range that allows for a range of proportions depending on the level of granularity intended by the speaker. The existential quantifiers require a meaning above 0, with a hypothetical intended proportion that includes the entire set, i.e., 'all.' The vague quantity expressions (excluding *most*)

11 Doran et al. (2009) originally set out to test whether scale structure (discrete or continuous) and boundedness (open or closed) affect the degree of implicature (Baker, Berends, Djalali, Doran, Larson, McNabb & Ward 2008) but ultimately found that expression type (divided into the categories numerals, quantifiers and modals, and gradable adjectives) was a more reliable factor.

12 Google Web 1T 5-Gram Database, for example, is an extremely large corpus but less accurate than others, such as the British National Corpus (Evert 2010; Michel, Shen, Aiden, Veres, Gray, Pickett, Hoiberg, Clancy, Norvig, Orwant et al. 2011; Biemann, Bildhauer, Evert, Goldhahn, Quasthoff, Schäfer, Simon, Swiezinski & Zesch 2013).

Quantifier	N-gram	<i>t</i> -score	Frequency
most	most if not all	369.83	138580
	most, but not all	284.45	81901
	most but not all	167.61	31118
	most, and perhaps all	23.87	581
some	some but not all	263.64	73541
	some if not all	91.92	10707
	some, and perhaps all	18.76	367
many	many if not all	116.75	14988
	many but not all	115.21	15560
	many, and perhaps all	14.91	231
a few	a few but not all	20.08	433
half	half, if not all	10.51	214
	half, but not all	6.09	186
	half if not all	-6.80	115
	half but not all	-13.88	151

Table 2 Collocations of a few quantifiers with *all*

denote a larger quantity, here arbitrarily represented as a proportion of more than half and including the entire set. Note that *most*'s range in the plot only approaches the proportion 1.0 although in some contexts it would be compatible with it (Ariel 2004).

If the range of posited communicated proportions given in Figure 2 is correct, then just considering the potential intended range of meaning of each quantifier would predict how compatible it would be with *all*. Put in other words, predicting what possible subsets each quantifier denotes predicts the degree to which the implicature 'not all' is likely to arise even without developing a more complex division between groups of quantifiers based on their shared semantic properties. The next step is to collect data about speakers' prior expectations of the range of potential proportions each of these quantifiers can be used to refer to, and if indeed these are similar to those hypothesized in Figure 2, then these prior expectations may very well be a crucial factor in the probability of a speaker to generate the scalar implicature 'not all.'

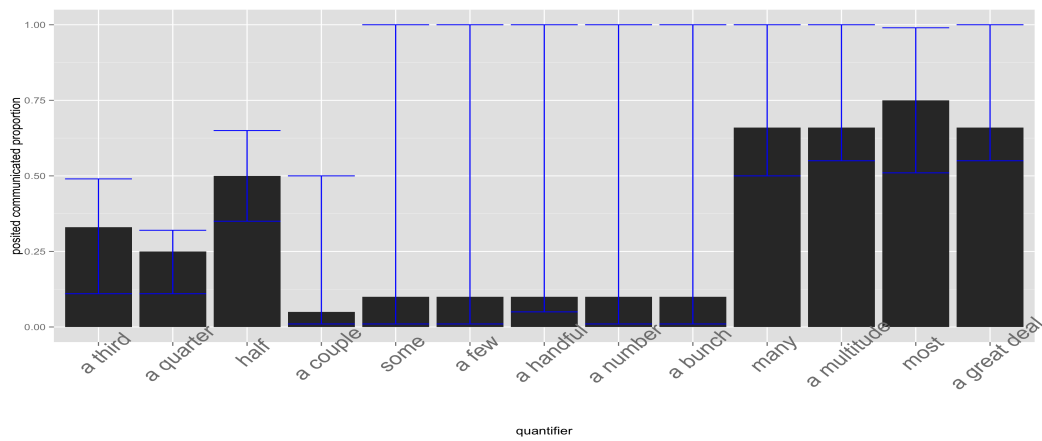


Figure 2 Posited communicated proportions for 13 quantifiers

6.3 *At most* and existential presupposition

Most of this section was dedicated to the interpretation of the compatibility of the 11 studied quantifiers with *all* when they were unmodified. The compatibility scores for unmodified quantifiers were varied in comparison with their compatibility scores when modified by *at most*. There, they were consistently incompatible with *all*, as expected. On the assumption that *at most* *Q* creates a downward-monotone environment, it should be compatible with *none* as well as with *some*, as already discussed in Section 4. Participants in these experiments, however, gave all quantifiers modified by *at most* high compatibility scores with *some* and low compatibility scores with *none*. Though very low, the compatibility scores with *none* were significantly higher than those for *all*.

The results from this study, then, point to an interpretation of *at most* *Q*, where *Q* stands for the 11 quantifiers studied in this experiment, that either includes an existential presupposition—consistent with Krifka’s (1999) observations—or some inference to that effect, which excludes the possibility of an empty set. The preferred inference under which the entailment that *none of the Q* is blocked given the assertion *at most Q* is reported in Sanford & Moxey 2004. They account for their results by appealing to the distinction between denial and affirmation. Denial is an often pragmatic property association with negation (Horn 2001). A listener that hears the assertion in (30) supposes that the speaker had reason to expect (that the listener had reason to expect) that Buffy had fought vampires last night. The assertion in (30) thus denies the expected state of affairs.

(30) Buffy didn’t fight any vampires last night.

Klima (1964) observes that tag questions point to whether a sentence is a negative or an affirmative sentence. The negative sentence in (31a), for example, takes *did she?* as a tag, whereas the affirmative sentence in (31b) takes *didn't she?* as a tag.

- (31) a. Buffy didn't fight any vampires last night, {did she / *didn't she}?
 b. Buffy fought a couple vampires last night, {*did she / didn't she}?

And so, if a positive tag as in (31a) points to sentence negation or a denial of a presupposed or expected state of affairs and a negative tag as in (31b) points to an affirmative sentence, then the well-formedness of *didn't she?* in (32) suggests that *at most Q* forms an assertion even though it's analytically monotone decreasing.

- (32) Buffy fought at most some of the vampires last night, {*did she / didn't she}?

In conclusion, despite the monotone-decreasing property of *at most Q*, which includes the empty set (i.e., *none of the N*), listeners reject this inference possibly because sentences with *at most Q* form assertions and not denials of expected states of affairs. When a speaker uses an assertion with *at most Q*, then the quantity denoted by *Q* has to be relevant to the exclusion of the empty set.

7 Conclusion

This paper presents a first, systematic investigation of the degree to which various quantifiers give rise to the implicature 'not all' using an utterance compatibility task with a modified Likert scale. Two accounts for the variation among quantifiers were proposed: (i) shared semantic properties among three coherent groups of quantifiers account for the degree they give rise to upper-bound implicature; or (ii) the likelihood of an implicature is a function of the scalar distance between the various quantifiers and 'all.' The predictions these two accounts make were discussed, charting the way to a future investigation of the heterogeneity of scalar implicature.

References

- Ariel, Mira. 2004. Most. *Language* 80(4). 658–706.
 Baker, Andrew. 2004. Japanese whisky culture: How it became both staple and status symbol. *The Telegraph*, November 6 .
 Baker, Rachel, Matthew Berends, Alex Djalali, Ryan Doran, Meredith Larson, Yaron McNabb & Gregory Ward. 2008. The effects of scale type and salience on the interpretation of scalar implicature. In *LSA Annual Meeting*, Linguistic Society of America.
 Barwise, Jon & Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4. 159–219.

- Biemann, Chris, Felix Bildhauer, Stefan Evert, Dirk Goldhahn, Uwe Quasthoff, Roland Schäfer, Johannes Simon, Leonard Swiezinski & Torsten Zesch. 2013. Scalable construction of high-quality web corpora. *Journal for Language Technology and Computational Linguistics* 28(2). 23–60.
- Bott, Lewis & Ira A. Noveck. 2004. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language* 51(3). 437–457.
- Christensen, Rune Haubo Bojesen. 2012. Ordinal Regression Models for Ordinal Data R, Package Version 2012.01-19.
- Cummins, Chris & Napoleon Katsos. 2010. Comparative and superlative quantifiers: Pragmatic effects of comparison type. *Journal of Semantics* 27(3). 271–305.
- Davies, Mark. 2008. The corpus of contemporary American English: 425 million words, 1990-present.
- Degen, Judith. 2015. Investigating the distribution of some (but not all) implicatures using corpora and web-based methods. *Semantics and Pragmatics* 8(11). 1–55.
- Doran, Ryan, Rachel M. Baker, Yaron McNabb, Meredith Larson & Gregory Ward. 2009. On the non-unified nature of scalar implicature: An empirical investigation. *International Review of Pragmatics* 1(2). 211–248.
- Doran, Ryan, Gregory Ward, Meredith Larson, Yaron McNabb & Rachel E. Baker. 2012. A novel experimental paradigm for distinguishing between ‘what is said’ and ‘what is implicated’. *Language* 88(1). 124–154.
- Evert, Stefan. 2010. Google web 1t 5-grams made easy (but not for the computer). In *North American Association for Computational Linguistics: Human Language Technologies (NAACL HLT) Sixth Web as Corpus Workshop*, 32–40. Association for Computational Linguistics.
- Gasdar, Gerald. 1979. *Pragmatics: Implicature, Presupposition, and Logical Form*. Academic Press New York.
- Geurts, Bart. 2006. Take ‘five’: The meaning and use of a number word. In Lilliane Tasmowski & Svetlana Vogeleeer (eds.), *Indefiniteness and Plurality*, Amsterdam: Benjamins.
- Geurts, Bart. 2010. *Quantity Implicatures*. Cambridge University Press.
- Geurts, Bart & Rick Nouwen. 2007. ‘At least’ et al.: The semantics of scalar modifiers. *Language* 83(3). 533–559.
- Guasti, Maria Teresa, Gennaro Chierchia, Stephen Crain, Francesca Foppolo, Andrea Gualmini & Luisa Meroni. 2005. Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes* 20(5). 667–696.
- Hackl, Martin. 2009. On the grammar and processing of proportional quantifiers: ‘most’ versus ‘more than half’. *Natural Language Semantics* 17. 63–98.
- Hansen, Nat & Emmanuel Chemla. 2013. Experimenting on contextualism. *Mind &*

- Language* 28(3). 286–321.
- Hirschberg, Julia Linn Bell. 1985. *A theory of scalar implicature*: University of Pennsylvania Ph.D. dissertation.
- Horn, Laurence. 1992. The said and the unsaid. In Chris Barker & David Dowty (eds.), *Semantics and Linguistic Theory (SALT) 2* (The Ohio State University Working Papers in Linguistics 40), 163–192. Columbus: Ohio State University.
- Horn, Laurence R. 1972. *On the Semantic Properties of Logical Operators in English*: UCLA Ph.D. dissertation.
- Horn, Laurence R. 2001. *A Natural History of Negation*. Stanford: CSLI Publications. Originally published 1989 by University of Chicago Press.
- Huang, Yi Ting, Elizabeth Spelke & Jesse Snedeker. 2013. What exactly do numbers mean? *Language Learning and Development* 9(2). 105–129.
- Kennedy, Christopher. 2013. A scalar semantics for scalar readings of number words. In Ivano Caponigro & Carlo Cecchetto (eds.), *From Grammar to Meaning: The Spontaneous Logicity of Language*, Cambridge University Press.
- Klima, Edward S. 1964. Negation in English. In Jerry A. Fodor & Jerrold J. Katz (eds.), *The Structure of Language*, 246–323. Englewood Cliffs, NJ: Prentice-Hall.
- Krifka, Manfred. 1999. At least some determiners aren't determiners. In K. Turner (ed.), *The Semantics/Pragmatics Interface from Different Points of View*, 257–291. Oxford: Elsevier Science.
- Lasnik, Peter. 1999. Pragmatic halos. *Language* 75(3). 522–551.
- Levinson, Stephen C. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT Press.
- McNabb, Yaron & Doris Penka. 2014. The interpretation of superlative modifiers and deontic modals: An experimental investigation. In Urtzi Etxeberria, Anamaria Fălăuș, Aritz Irurtzun & Bryan Leferman (eds.), *Sinn und Bedeutung (SuB) 18*, 271–288. Bayonne and Vitoria-Gasteiz.
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant et al. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014). 176–182.
- Musolino, Julien. 2004. The semantics and acquisition of number words: Integrating linguistic and developmental perspectives. *Cognition* 93(1). 1–41.
- Pouscoulous, Nausicaa, Ira A Noveck, Guy Politzer & Anne Bastide. 2007. A developmental investigation of processing costs in implicature production. *Language Acquisition* 14(4). 347–375.
- R Development Core Team. 2006. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.
- Sanford, Anthony J. & Linda M. Moxey. 2004. Exploring quantifiers: Pragmatics meets the psychology of comprehension. In Ira A. Noveck & Dan Sperber (eds.),

- Experimental Pragmatics*, 116–137. Palgrave Macmillan.
- Sauerland, Uli & Penka Stateva. 2007. Scalar vs. epistemic vagueness: evidence from approximators. In M. Gibson & T. Friedman (eds.), *Semantics and Linguistic Theory (SALT) 27*, 228–245. New York: CLC Publications.
- Solt, Stephanie. 2009. *The Semantics of Adjectives of Quantity*: The City University of New York Ph.D. dissertation.
- van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina & Bart Geurts. 2014. Scalar diversity. *Journal of Semantics* DOI: 10.1093/jos/ffu017.

Yaron McNabb
Department of Languages, Literature and Communication
Utrecht Institute of Linguistics (UiL)
Utrecht University
Trans 10, Utrecht
Netherlands
y.mcnabb@uu.nl