4

MOST MEANINGS ARE SUPERLATIVE☆

Hadas Kotek, Yasutada Sudo, Edwin Howard and Martin Hackl

ABSTRACT

Recent research on the semantics of quantificational expressions has taken on the task of relating the truth-conditional import of quantifiers with their impact on verification procedures. In particular, the semantic analysis and verification procedures associated with the proportional quantifier *most* have been studied in a variety of ways in an effort to reveal the correct semantic analysis of *most*. This chapter presents new experimental evidence in favor of a treatment that analyzes *most* uniformly as a superlative construction built in the syntax, rather than a quantificational determiner with rich internal semantics that does not, however, interface with syntax any more than other lexical determiners do. The evidence comes in the form of a latent superlative reading of sentences with *most* in subject position, like *most of the dots are blue*. We show, using three different

☆Yasutada Sudo, Hadas Kotek and Martin Hackl designed Experiments 1 and 2; Hadas Kotek and Martin Hackl designed Experiment 3; Yasutada Sudo analyzed the data. Hadas Kotek proposed the analysis and Edwin Howard and Martin Hackl developed it. Hadas Kotek and Martin Hackl with input from Edwin Howard and Yasutada Sudo, wrote the chapter.

Experiments at the Interfaces
Syntax and Semantics, Volume 37

ISSN: 0092-4563/doi:10.1108/S0092-4563(2011)0000037008

experiments involving verification tasks, that *most* is ambiguous in this environment between a dominant proportional reading and a latent superlative reading.

1. INTRODUCTION

The study of natural language quantification is a central topic in syntax and semantics that provides crucial insights into the complexity of language. The phenomena that researchers have traditionally concentrated on when characterizing quantificational expressions are the entailment patterns associated with specific quantifiers and their combinatorial properties. However, as is well known, these properties alone under-determine the correct semantic analysis of natural language quantification. In an effort to narrow the space of possible analyses, researchers have begun to consider new empirical domains. One area of expansion concerns the investigation of verification procedures associated with particular quantificational expressions. In particular, the verification procedures associated with the proportional quantifier *most* have been studied in a variety of ways in an effort to reveal the correct semantic analysis of *most* (e.g., Hackl, 2009; Lidz, Pietroski, Hunter & Halberda, 2011; Solt, 2011).

It is well-known that *most* occurs both as a proportional determiner, in which case it can be paraphrased using *more than half*, (1)a, and as the spellout of the superlative morpheme -*est*, a reading that appears when *most* combines with certain kinds of degree predicates (e.g., *most expensive car*). In this case it seems to always be accompanied by the definite article. One instance of the superlative use of *the most* arises when -*est* combines with the gradable predicate *many*/*much* to give rise to the amount superlative construction, (1)b (Bresnan, 1973).

(1) a. John talked to most students. **proportional**
$\approx$ John talked to more than half of the students
b. John talked to the most students. **superlative**
$\approx$ John talked to more students than anybody else

There is a debate in the literature as to whether or not the two uses of *most* in (1) should be related to one another. Under the standard Generalized Quantifier Theory (GQT, Barwise & Cooper, 1981) view, bare *most* and superlative *most* are separate lexical items that happen to share a phonological exponent. That is, (1)a–b will be analyzed as a lexical ambiguity. An alternative approach is to view (1)a–b as a structural ambiguity. Under this view, there is a common lexical source that gives rise to different truth-conditions based on the logical form of the sentence it appears in (e.g., Hackl, 2009; Yabushita, 1999).[1]

[1] See Szabolcsi (2010) for extensive review of the history of this debate.

This chapter contributes to the debate over the correct analysis of *most*. Specifically, we present experimental evidence that bare *most* itself, when it appears in subject position and seems to unambiguously serve as a quantificational determiner, is ambiguous between a proportional and a superlative reading. To the best of our knowledge, this fact has not previously been noticed in the literature.

We use this observation to argue for two points, one methodological and one theoretical. The methodological point is that verification experiments can reveal aspects of meaning that native speaker intuitions about the meaning of expressions might miss. In particular, we suggest that the ambiguity of *most* in subject position has gone unnoticed because the two readings are not equally accessible: the proportional reading is dominant and masks the existence of the superlative reading. As a consequence, probing for native speaker intuitions regarding entailment patterns associated with *most* may not be a feasible approach to detecting the ambiguity. Instead, specific tasks demands, such as those that obtain in our sentence verification experiments, must be in place in order to bring the latent superlative reading of *most* to the foreground.

The theoretical significance of our observation is that it provides compelling evidence for a decompositional analysis of *most* under which it is uniformly analyzed as a superlative construction. To account for the observation that bare *most* can give rise to superlative truth-conditions, the lexical ambiguity view would have to postulate yet another layer of lexical ambiguity: $most_{PROP}$, $most_{SUP}$, and *the most*, as well as ad hoc rules for when $most_{PROP}$ and $most_{SUP}$ should be used. The structural ambiguity view, however, can capture the ambiguity very naturally, as we will show.

The chapter is structured as follows. Section 2 briefly reviews the debate between the approach that views *most* as a lexical primitive and the approach which views *most* as a complex superlative construction. In Section 3 we present novel data from three different experiments which support the view that *most* is uniformly a superlative construction. Specifically, we show that under certain circumstances speakers can access a latent superlative interpretation of *most* in subject position, in addition to the expected proportional meaning. We compare our results with the findings of a related study in Lidz et al. (2011) in Section 4 and argue that the seemingly conflicting results of that study are a result of quite different task demands. We, then, develop an account of the ambiguity within a decompositional analysis of *most* which exploits the mechanism of association with focus. Conclusions are drawn in Section 5.

2. PROPORTIONAL AND SUPERLATIVE *MOST*

As we have seen in (1)a–b, *most* is ambiguous between a proportional reading and a superlative reading when it appears in object position. The availability of these two readings correlates with the presence or absence of the definite article: bare *most* gives rise to a proportional reading, while *the most* produces a superlative reading.

Under the lexical ambiguity view, the proportional reading of *most* is analyzed as in (2): bare *most* is a quantificational determiner that combines with two set-denoting expressions A and B and yields true only if the intersection of the sets A and B is more numerous than the intersection of the set A and the complement of the set B. This holds if more than half of the As are Bs.

(2) a. John talked to most students.
b. $[\![\text{most}]\!](A)(B) = 1$ iff $|A \cap B| > |A - B|$
c. $[\![\text{John talked to most students}]\!] = 1$ iff $|\{x : x \text{ is a student}\} \cap \{x : \text{John talked to } x\}| > |\{x : x \text{ is a student}\} - \{x : \text{John talked to } x\}|$

Following Heim (1985), Szabolcsi (1986), the superlative reading can be analyzed as a construction that involves degree quantification, (3), with the superlative morpheme *-est* denoting a degree quantifier that is restricted by a comparison class C, which, in the case of (3)a contains contextually salient alternatives to John. (3)a is true only if there is a plurality of students that John talked to that is more numerous than any plurality of students talked to by any contextually salient individual different from John.[2]

(3) a. John talked to the most students.
b. $[\![\text{John talked to the most students}]\!] = 1$ iff $\exists d \exists X[\text{students}(X) \,\&\, \text{John talked to } X \,\&\, |X| \geq d \,\&\, \forall y \in C[y \neq \text{John} \rightarrow \neg\exists Y[\text{students}(Y) \,\&\, y \text{ talked to } Y \,\&\, |Y| \geq d]]$

On the structural ambiguity view, both the proportional and the superlative readings of *most* are analyzed as superlative constructions. The proposal in Hackl (2009) analyzes the superlative reading of *most* as in (3) above. The proportional reading, on the other hand, is analyzed as in (4). (4)b is parallel to (3)b except that (i) the comparison class C is assumed to be the set of students rather than the set of people who talked to students and (ii) nonidentity holds between any two alternatives in C if they are nonoverlapping pluralities of students. We use $\perp$ to represent the no-overlap relation, which replaces the nonidentity relation $\neq$ of (3)b. (4)a is true just in case there is a plurality of students that John talked to that is more numerous than all its complements — that is, the student pluralities that John did not talk to. (4)a expresses

[2]We ignore in our formulas the difference between sets and pluralities whenever the distinction is immaterial to the discussion. For instance, the cardinality function $|.|$ is used for both sets and pluralities.

proportional truth-conditions. It is true only if John talked to more than half of the students.[3]

(4) a. John talked to most students.
b. $[\![$John talked to most students$]\!] = 1$ iff $\exists d \exists X$[students(X) & John talked to X $\&|X| \geq d$ & $\forall Y \in C$[students(Y)&$Y \perp X \rightarrow |Y| < d$]]

As they stand neither analysis imposes any constraints on the syntactic distribution of the proportional and superlative readings. Specifically, we expect *the most* and bare *most* to occur freely in subject position. This does not seem to be empirically adequate, however. Rather it seems that bare *most* in a partitive frame appears to be the most natural way to express a statement with *most* in subject position,[4] (5)a, while its counterpart *the most* is simply ungrammatical, (5)b, and the status of (5)c is at least degraded relative to (5)a.[5]

(5) a. Most of the students talked to John.
b. *The most of the students talked to John.
c. ??The most students talked to John.

Given that the grammatical status of *the most* in subject position is unexpectedly degraded compared to the status of *the most* in object position, we are interested in whether superlative truth-conditions are expressible with *most* in subject position at all. Specifically, we are interested in knowing whether a sentence like (5)a can ever be used to express superlative truth-conditions, or only proportional ones. Based on its lack of a definite article, we would expect that it cannot have a superlative reading and indeed informal probing of native speaker intuitions about the truth-conditions of sentences like (5)a seems to support this. However, we will present evidence that matters are

---

[3]To ensure that (4)b does not express "absolute" truth-conditions, which would be paraphrasable by *all the students*, we need to assume that C contains at least two distinct elements. Hackl (2009) argues that this is a presupposition of the superlative operator.

[4]See Matthewson (2001) for a discussion of *most* with and without the partitive construction.

[5]At least some speakers find (5)c entirely ungrammatical while others find that it can be uttered given sufficient supporting context, as in (i). We thank Noah Constant (p.c.) for bringing this to our attention. In general, however, it seems that the status of (5)c is diminished compared to (5)a.

(i) Ten people arrived at 8 am, twenty arrived at 9 am, and the most (people) arrived at 10 am.

It has been argued that the reason for the ungrammaticality of *the most* in subject position is the restriction on left-branch extraction in English. This sets English apart from German, where there is no such restriction. Sentences containing *die meisten*, such as the equivalent of (5) can consequently have either a proportional (= more than half) or a superlative (= more than any other color) interpretation. Both readings are generally equally accessible to speakers:

(ii) Die meisten Kreise sind blau.

more complex than that. In particular, we show that in certain circumstances sentences like (5)a do allow for a superlative interpretation. This fact is unexpected under both approaches to *most*. They differ, however, in how they might go about capturing the new reading. The only way that the lexical ambiguity view of *most* can capture this reading is to proliferate lexical ambiguity. By contrast, we show that a structural ambiguity approach, which analyzes *most* uniformly as a superlative construction, can capture the ambiguity by simply adjusting the comparison class argument of the superlative operator.

3. VERIFYING *MOST*

In this section we present evidence from three sentence verification experiments that speakers can access a superlative reading of *most* in subject position. We first motivate the use of verification as a means to explore the semantics of *most*, looking both at the specific components of a decompositional analysis and how manipulations of an external scene might affect verification given these components. Next, we describe the logic underlying the experimental material and lay out predictions. Finally, we present the specific methodologies and findings of our experiments.

In using verification tasks to study the semantics of *most*, we follow a tradition in psycholinguistics that is based on the hypothesis that there is a systematic relation between the format in which the truth-conditions of a certain sentence are stated and how speakers verify that sentence, and that observations from one component can shed light on the structure of the other component (Carpenter & Just, 1975; Clark & Chase, 1972, among others; see Tanenhaus, Carroll, & Bever, 1976 for a critical review). A recent statement of this hypothesis is given in Lidz et al. (2011), in the form of the Interface Transparency Thesis (ITT).

(6) **Interface Transparency Thesis** (ITT)
The verification procedures employed in understanding a declarative sentence are biased toward algorithms that directly compute the relations and operations expressed by the semantic representation of that sentence.

Given the ITT, we expect that variation in scenes, with regard to which sentences are verified, can have systematic effects on the verification process. More precisely, the verification process should be sensitive to variation in the scenes if the variation targets information that the verification algorithms rely on.

We will be using verification tasks to probe for the existence of a superlative reading for sentences like *most of the dots are blue*, in which *most* occurs in subject position. Following Lidz et al. (2011), our strategy is to vary the number colors (two or three) in the dot arrays relative to which participants judge our target sentences in such a way that verification procedures should be affected if the decompositional analysis of *most* is correct.[6]

First, consider how the meaning of (7)a would be represented under the decompositional analysis, (7)b.

(7) a. Most of the dots are blue.

b. $[[\text{-est many dots}] \text{ are blue}] = 1 \text{ iff } \exists d \exists X[\text{dots}(X) \wedge \text{blue}(X) \wedge |X| \geq d \;\&\; \forall Y \in C\; [\text{dots}(Y) \;\&\; Y \perp X \rightarrow |Y| < d]]$

c. **Proportional truth-conditions**
$|\text{blue dots}| > |\text{non-blue dots}|$

d. **Superlative truth-conditions**
For each non-blue color Z, $|\text{blue dots}| > |\text{Z dots}|$

As mentioned in the previous section, the truth-conditions that (7)a is predicted to express under the decompositional analysis depend on the specific content of the comparison class, C. If C is identified with the extension of the plural NP *dots*, which, by assumption, is closed under i-sum formation, proportional truth-conditions result, (7)c. This is because all dot pluralities different from the blue dots — whether they are homogenous in color or not — need to be less numerous than the blue dots for the sentence to be true. Thus, on this setting of C, (7)b reduces to (7)c. If C is further constrained so that only homogenously colored dot pluralities are included, different truth-conditions result. For (7)a to be true under this setting of C it is sufficient that the blue dots outnumber each of the homogenously colored non-blue dot pluralities separately but it is no longer required that the blue dots outnumber the non-blue dots as a whole, (7)d. Hence there is no requirement that more than half of the dots are blue. In other words, if the meaning of (7)a is given as in (7)b and if C can be either a partial closure of dot pluralities (homogenous) or the full closure (nonhomogenous) then we expect (7)a to be ambiguous between (7)c and (7)d.

Next, consider under what circumstances this ambiguity could be detected. Clearly, if only two colors are used in an array there cannot be a difference between (7)c and (7)d since the proportional reading of *most* and the superlative reading of *most* predict the same truth-conditions for two-colored dot arrays. However, the two possible readings come apart in dot arrays that contain more than two colors. That is, the proportional reading of *most* entails the superlative reading of *most*, but the superlative reading does not entail the proportional reading.[7]

(8) a. $most_{\text{PROP}} \Rightarrow most_{\text{SUP}}$

b. $most_{\text{SUP}} \nRightarrow most_{\text{PROP}}$

[6]The experimental procedure used in Lidz et al. (2011) is quite different from the ones we use in our experiments and we suspect that this is the main reason why our results are rather different from theirs. See Section 4 for discussion.

[7]The two readings are in principle independent of each other, as Hackl (2009) shows.

Whenever the proportional reading of *most* is true, as is the case in the three-color array in Figure 1, the superlative reading will also be true: if the blue dots outnumber the set of non-blue dots as a whole, then they necessarily also outnumber each homogeneously colored subset of the non-blue dots individually.

However, the superlative reading of *most* can be true while the proportional reading is false, as is the case, for example, in Figure 2.

For a dot array as in Figure 2, which contains 7 blue dots, 4 red dots, and 4 yellow dots, the two possible verification strategies of the sentence *most of the dots are blue* make different predictions. Under a proportional reading of *most*, the number of blue dots, 7, is compared with the number of non-blue dots, 8, resulting in a false statement. Under a superlative reading of *most*, the number of blue dots, 7, is compared with the number of red dots, 4, and then separately with the number of yellow dots, 4. The blue dots outnumber each of the two homogeneously colored subsets of dots and consequently the statement should be judged true by speakers who verify it according to these truth-conditions. We will refer to this state of affairs as the "superlative" condition: the context in which we predict different results depending on whether or not speakers can access superlative truth-conditions for *most*.

Experiments 1–3 below take advantage of the prediction of the decompositional analysis of *most* that arrays like the one in Figure 2 can be verified in two

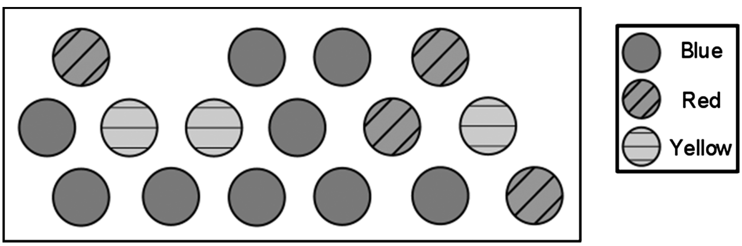


Figure 1: *Proportional* reading of *most* entails the *superlative* reading.

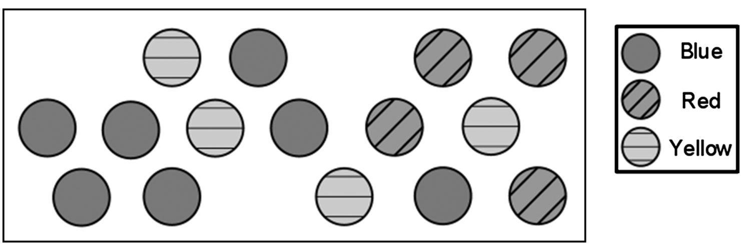


Figure 2: *Most of the dots are blue* only true under superlative reading.

conflicting ways. We show, more specifically, that both strategies are in principle available to speakers of English, although verification according to proportional truth-conditions is preferred.

In order to ensure that the experimental manipulation of the "superlative" condition is the source of any results obtained in our studies, as opposed to outside pressures that are unrelated to the semantic representation of *most*, we use the proportional determiner *more than half*, (9), as a control. According to the entry in (9), *more than half* should be analyzed as a comparative determiner, incorporating the notion of *half*.

(9) $[\![\text{more than half}]\!](A)(B) = 1 \text{ iff} |A \cap B| > \frac{1}{2}|A|$

Under the lexical ambiguity view of *most*, it is predicted that statements like *most of the dots are blue* and *more than half of the dots are blue* are true in exactly the same cases: whenever the blue dots outnumber the non-blue dots. This reflects the expected proportional reading of (bare) *most* and of *more than half* in this context. Under (9), there is no reason to expect that *more than half* will ever be verified according to superlative truth-conditions. Therefore, if we find that *most* and *more than half* show the same level of (in)sensitivity to the "superlative" condition, we will not be able to conclude that *most* has a semantic representation that allows it to be verified according to superlative truth-conditions. Rather, any sensitivity to the "superlative" condition will have to be attributed to outside factors governing the behavior of participants in our studies. If, on the other hand, we find that *most* is sensitive to the "superlative" condition but *more than half* is not, then we can conclude that the source of the difference between the two determiners is in their semantic representations and, following the ITT, in the verification strategies that are available to the two determiners.

The studies below will thus cross two factors, each with two levels: Determiner ("most" and "more than half") and Color (2C and 3C). Within the 3C conditions, trials that are true only under superlative truth-conditions but not under proportional truth-conditions, as in Figure 2 above, are the crucial test cases that have the potential to distinguish between *most* and *more than half*. Cases as in Figure 1, which are true under both proportional and superlative truth-conditions, cannot teach us whether or not participants have access to genuine superlative truth-conditions. We will refer to cases as in Figure 1 as "Yes_{PROP}" cases, since they are true under proportional truth-conditions; we will refer to cases as in Figure 2 as "No_{PROP}" cases, because they are false under proportional truth-conditions. This coding system will allow for a straightforward comparison with *more than half*, whose behavior we expect to track proportional truth-conditions only. Table 1 summarizes the two possible judgment patterns that our experiments will try to detect. For each of the four conditions that our experiments can contain, we give one example of a blue-to-non-blue dot ratio.

For both "Yes$_{\text{PROP}}$" conditions, proportional truth-conditions and superlative truth-conditions make the same predictions: the blue dots outnumber the non-blue dots as a whole, and consequently they also outnumber any subset of the non-blue dots. Therefore, we expect all of the participants in our experiments to answer "Yes" or "True" in these conditions regardless of whether they are verifying *most* trials or *more than half* trials and regardless of which truth-conditions they use to verify the trials with. Likewise in "2C-No$_{\text{PROP}}$" trials, proportional truth-conditions and superlative truth-conditions make the same predictions: the blue dots do not outnumber the only available set of non-blue dots. Therefore, we expect participants to always answer "No" or "False" in these trials. The critical difference between proportional and superlative verification strategies becomes clear in "3C-No$_{\text{PROP}}$" trials. According to proportional truth-conditions, participants will compare, for example, 8 blue dots to 9 non-blue dots and will answer "No" or "False." According to superlative truth-conditions, however, participants will compare 8 blue dots to 5 red dots and 4 yellow dots and will answer "Yes" or "True." Our experiments will thus take *more than half* as a baseline, with an expected behavior as in the left column in Table 1, and compare it with *most*, which could have an expected behavior as in the left column but also as in the right column.

Figure 3 demonstrates visually the effects of the two possible verification strategies of *most* statements in "No$_{\text{PROP}}$" trials. We use *more than half* as a baseline against which *most* can be compared. If participants always verify *most* statements according to proportional truth-conditions, then the Yes/No response rates for those trials should resemble the response rates observed

Table 1: Possible judgment patterns for *more than half* and *most*.

	Proportional truth-conditions	*Superlative truth-conditions*
2C-Yes$_{\text{PROP}}$ (9:8)	Yes	Yes
3C-Yes$_{\text{PROP}}$ (9:4:4)	Yes	Yes
2C-No$_{\text{PROP}}$ (8:9)	No	No
3C-No$_{\text{PROP}}$ (8:5:4)	No	**Yes**

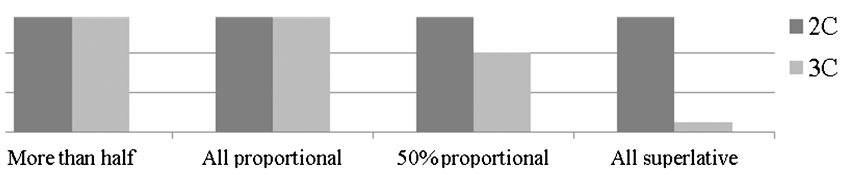


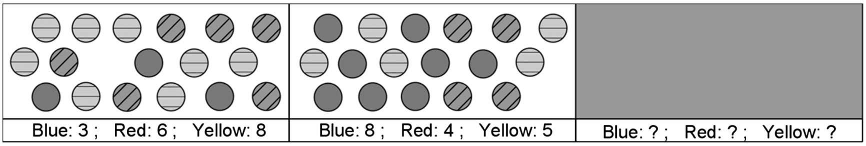
Figure 3: Projected response rates for "No$_{\text{PROP}}$" trials.

for *more than half*. If, for example, participants verify according to proportional truth-conditions half of the time and according to superlative truth-conditions the rest of the time, then we expect that the bar which represents No-responses to 2-Color trials will not be affected but that the bar which represents No-responses to 3-Color trials will be lower, representing an increase in Yes-responses to this condition. If participants use a superlative verification strategy all of the time, we expect again that the bar which represents responses to 2-Color trials will not be affected but that the bar which represents responses to 3-Color trials will drop to zero, signifying that participants responded "Yes" to this condition all of the time.

3.1. Experiment 1: Covered Box

The purpose of Experiment 1 is to determine whether the superlative reading of *most* in subject position is available to native speakers of English. To this end, we employ Snedeker's "covered box" task (Huang, Spelke, & Snedeker, ms.; Pearson, Khan, & Snedeker, 2010). The covered box task is a variant of the picture matching task in which participants are given a sentence and a set of pictures, and are asked to choose the picture that best matches the situation that the sentence describes. Unlike in the normal picture matching task, however, one of the pictures is a covered box; in our experiment, participants were told to choose the covered box only if they think the other pictures do not match the sentence. Participants are taught in the practice phase of the experiment that exactly one of the possible three pictures matches a given sentence. That is, if one of the uncovered pictures matches the sentence, the picture under the covered box does not match it, and if the two uncovered pictures do not match the sentence, then the picture under the covered box does. After the practice phase participants are no longer allowed to see what is hidden in the covered box.

The covered box method is especially suited for detecting less preferred interpretations of ambiguous sentences: for a sentence like *most of the dots are blue*, we noted above that a true "proportional" picture — that is, one in which the blue dots outnumber the non-blue dots as a whole — entails that the superlative truth-conditions of *most* will also be met. Therefore, presenting a "proportional" picture will not yield conclusive evidence about whether speakers use a proportional verification strategy or a superlative one. The critical case involves pictures which are only true under the superlative reading but not under the proportional reading. If the "superlative" picture is the only one which is presented then we are able to learn whether speakers can verify the sentence according to superlative truth-conditions: if they are, then the sentence *most of the dots are blue* is true and speakers should choose the superlative picture. If they are unable to verify according to superlative truth-conditions, then the uncovered picture does not match the sentence and they should choose the covered box.



(a) False picture (b) Superlative picture (c) Covered box

Figure 4: A covered box trial in Experiment 1.

3.1.1. Methods and Materials

Each trial in Experiment 1 consists of a sentence and a set of two pictures and a covered box. Each non-box picture contains 15–18 dots in three colors, blue, red, and yellow. The number of blue, red, and yellow dots that each picture contains is listed under the picture.[8] These pictures were presented in combinations with a sentence that either employed *most* or *more than half*, where Determiner was a between-subject factor — that is, participants saw only *most* statements like (10) or only *more than half* statements like (11). An example target trial looks as in Figure 4. The target trials that these statements describe are identical: the exact same pictures were used for both conditions, changing only the sentence matched with the pictures. The distractor trials were all identical in both conditions.

(10) Most of the dots are blue

(11) More than half of the dots are blue

The picture in (a) is a distractor that is false under both possible readings of *most*. The picture in (b), which is similar to Figure 4 above, is true under superlative truth-conditions but, crucially, not under proportional truth-conditions. Therefore, if speakers can access a superlative reading of *most*, (b) can be chosen. If the superlative reading is unavailable, the covered box (c) will be chosen. Speakers who verify *more than half* statements, by comparison, only have proportional truth-conditions available for this determiner. Therefore, the pictures in (a) and (b) should not be chosen; rather, speakers should always choose the covered box.

The dependent measure is the rate of (a), (b), and (c) answers for the two determiners, with *more than half* acting as a baseline. For *more than half* we expect that (a) and (b) should both not be chosen at all. If a superlative reading for *most* is available, the number of (b) answers will be significantly higher than

[8] A previous version of this experiment which did not list the numbers suggested that participants did not count the dots by themselves but rather provided estimated responses which proved to be inaccurate across all experimental conditions.

the number of (a) answers, showing that (b) was not chosen by error. However, it is still possible that (c) is preferred to (b) in general, depending on how easy it is to access the truth-conditions that would make (b) a viable choice.

We posted surveys on Amazon Mechanical Turk.[9] Participants were paid at the rate of $0.10 for their participation. They were asked to indicate their native language, but payment was not contingent on their response.

The surveys had 4 target items and 14 filler items. The distractors had unambiguously correct answers. The sentences for 2 of the filler items were *most/more than half of the dots are blue*; 6 contained the determiner *many*, 2 contained the proportional determiner *more than 50%*, and 4 contained the proportional determiners: *more than 1/3, more than 2/3, more than 1/4, more than 3/4.* In half of the filler trials, the correct answer was the covered box and in the other half, the correct answer was one of the overt pictures. This allowed us to determine whether participants select the covered box at all when the overt pictures are both incorrect, and employ an exclusion criterion that was partially independent of our target items for participants whose error rate was too high. The items were presented in one of the two pseudorandomized orders where each pair of target items was separated by at least one filler item, and the first two items were not target items.

3.1.2. Results

Forty-two monolingual native speakers of English participated in this study. Twenty-three subjects were shown the "most" condition and 19 were shown the "more than half" condition. In addition to these participants, 6 subjects were excluded from the analysis because of low accuracy rates ($<85\%$ for all trials); 5 nonnative speakers and 16 multi-lingual speakers were also excluded from the analysis.

The results show that the covered box was chosen 67.4% of the time and the superlative picture was chosen 32.6% of the time in the target items in the "most" condition. The difference between the preferences of the two pictures is significant by a Wilcoxon signed rank test ($W = 172.5$, $Z = 2.063$, $p < 0.05$). The false distractor pictures in the target trials were never chosen.

We also observe that 10 of the participants always selected the covered box in the target trials. The remaining 13 participants in the "most" condition selected the superlative picture at least some of the time: 5 participants chose it once, 2 participants chose it twice, 3 participants chose it three times, and 3 participants chose it four times.

In the "more than half" condition, the superlative picture was never chosen. Rather, the covered box was chosen 100% of the time. The superlative answer is significantly more frequent in the "most" condition than in the "more than half" condition (Mann–Whitney U test: $U = 342$, $Z = 3.816$, $p < 0.001$).

[9]https://www.mturk.com/mturk/welcome

3.1.3. Discussion

The results suggest that the superlative reading exists for *most* but not for *more than half*. The fact that the superlative picture was chosen a third of the time for *most* but never for *more than half* suggests that the observed patterns stem from the different semantic representations of *most* and *more than half* rather than from the experimental manipulation. A second important finding of Experiment 1 is that although the superlative reading is in principle available, it seems less preferred than the proportional reading of *most*. That is, the superlative picture is only chosen by a subset of the participants in Experiment 1 while others never choose it. Additionally, looking at individual participants who can access the superlative reading, it seems that they choose the superlative picture only some of the time, and the covered box at other times. This suggests that the proportional reading is dominant and available to all speakers of English, while the superlative reading is latent and available only to some speakers. Alternatively, it is possible that the superlative reading is available to all speakers but it is more difficult to access for some speakers than for others.

3.2. Experiment 2: Ratings

Having shown that the superlative reading of *most* can in principle be accessed by English speakers but that it is in general less accessible than the proportional reading, Experiment 2 introduces a new layer of complexity that was absent from Experiment 1: the Color manipulation. Pictures in this experiment contain either two colors or three colors. In addition, we use a geometric arrangement of dots that makes it unnecessary to specify the number of dots in each color in the arrays, and the Determiner manipulation is within-subject, allowing for a direct comparison between *most* and *more than half*. Instead of the covered box method, Experiment 2 is a simple rating task.

3.2.1. Methods and Materials

In each trial in Experiment 2, participants are asked how confident they are in answering “Yes” or “No” to a question about a picture which contains 15–19 figures (circles, triangles or squares; each picture contains just one type of shape) in either two or three colors out of blue, gray, and yellow.[10] They are given a choice of 8 radio buttons which are labeled NO on the far left and YES on the far right. They are encouraged to choose buttons in the extreme edges if they are confident in their answer and buttons in the middle range if they are less confident in their answer. Target trials ask participants one of the two questions in (12)–(13). The figures in the pictures are always arranged in three

[10]Unless otherwise specified, we will always use “dots” to refer to the figures in the arrays, and “blue dots” to refer to the target set.

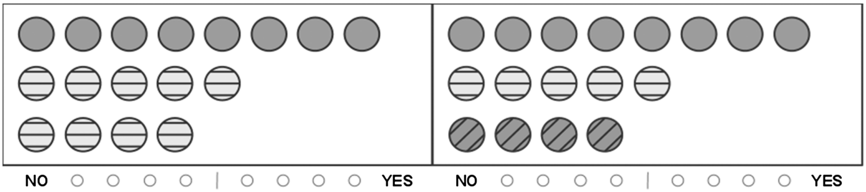


Figure 5: Sample 2-Color and 3-Color pictures in Experiment 2.

rows, as in Figure 5. This arrangement allows for a simple verification strategy according to superlative truth-conditions but also for a relatively simple verification strategy according to proportional truth-conditions.

(12) Are most of the dots blue?

(13) Are more than half of the dots blue?

3-Color target trials are true only under the superlative reading of *most* but not under the proportional reading. There are 6 target trials for each of the two Color conditions for each of the two target determiners. Determiner was treated as a within-subject factor in this experiment, giving a total of 24 target items. The same pictures were used with *most* trials and with *more than half* trials, so that each target picture was used twice in Experiment 2. The order of the pictures was controlled so that identical pictures did not appear one right after the other. Moreover, between two target trials with identical pictures, there is always at least one target trial with a different picture. There were 72 filler trials in this experiment, 12 of which were *most* trials and 12 of which were *more than half* trials. These trials were paired with pictures that are clearly true under the proportional reading of *most* and with pictures that are clearly false under both readings of *most*. Two other distractor determiners were used: *more than 1/3* and *many*, each paired with 24 pictures. In half of the filler items the answer was in the No-range, and in half it was in the Yes-range. The items were presented in two pseudorandomized ordered such that two trials with the same determiner are separated by at least one trial with a different determiner, and the first six items are all filler items.

The survey was posted on Amazon Mechanical Turk. Participants were paid at the rate of \$0.30 for their participation. They were asked to indicate their native language, but payment was not contingent on their response.

3.2.2. Results

Thirty-four subjects were recruited on Amazon Mechanical Turk, among whom 5 nonnative speakers of English, 3 subjects whose accuracy rate for the filler items is less than 80%, and 1 subject who did not answer one target item were excluded from the analysis. The mean ratings and by-subject z-scores of

Table 2: By condition mean ratings and z-scores for *most* and *more than half* in Experiment 2.

Determiner	Color	Mean rating	Mean z-score
Most	2C	2.508 ($SE = 0.280$)	−0.506 ($SE = 0.070$)
Most	3C	3.908 ($SE = 0.432$)	−0.035 ($SE = 0.114$)
More than half	2C	1.975 ($SE = 0.228$)	−0.682 ($SE = 0.054$)
More than half	3C	2.517 ($SE = 0.330$)	−0.492 ($SE = 0.088$)

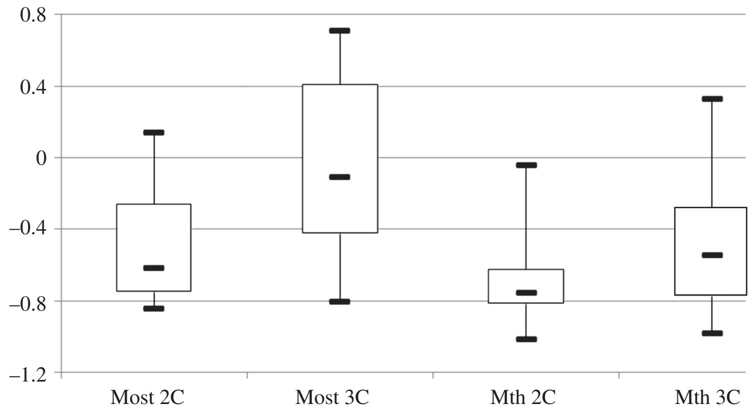


Figure 6: z-scores for *most* and *more than half* across Color conditions in Experiment 2.

the target items sorted by condition are given in Table 2 (standard errors in parentheses). The z-scores are also given visually in Figure 6.

We fit a mixed linear model to the z-scores. The model includes random intercepts for both subjects and items, and a by subject random slope for the effect of Determiner.[11] Log-likelihood ratio tests show that the by subject random slope for Determiner significantly improves the model fit ($\chi^2(2) = 22.419$, $p < 0.001$), while the inclusion of the by subject random slope for Color does not ($\chi^2(3) = 4.162$, $p = 0.245$). By-items random slopes

[11]The model was fit using the *lmer* function in R in the *lme4* package (using the R code below). The predictor variables are coded as follows:

(i)	Determiner:	Most = 0.5	More than half = −0.5
	Color:	3C = 0.5	2C = −0.5
(ii)	lmer(z-score ~ Determiner*Color+(1+Determiner∣SubjectID)+(1∣ItemID))		

Table 3: Summary of the fixed effects.

Predictor	Coefficient	Standard error	t
Intercept	−0.429	0.057	−7.571
Determiner	0.136	0.101	3.121
Color	0.331	0.055	6.027
Determiner × Color	0.281	0.110	2.563

Table 4: Summary of the random effects.

Group	Random effect	s^2	Standard deviation
Item	Intercept	0.002	0.035
Subject	Intercept	0.049	0.222
	Determiner	0.145	0.381

are also superfluous and are not included in the model (the results of the log-likelihood ratio tests are omitted here).

The fixed and random effects are summarized in Tables 3 and 4 ($N = 480$, log-likelihood $= -452.9$). The correlations among the fixed effects are all within ± 0.240, and the correlation between the random effects for subjects is 0.327 .

A log-likelihood ratio test with the model without the fixed interaction effect shows that it significantly improves the model fit ($\chi^2(1) = 6.390$, $p < 0.05$).

3.2.3. Discussion

We find that the interaction between Determiner and Color is significant in Experiment 2. Specifically, we find that for *more than half*, participants give both color conditions similarly low ratings within the No-range; for *most*, we find that participants rate the two color conditions differently: the 3-Color condition is rated significantly higher than the 2-Color condition, although all average ratings remain within the No-Range. Recall that the 3-Color-No$_{\text{PROP}}$ condition for *most* corresponds to the "superlative" condition. If participants are able to verify *most* according to superlative truth-conditions, in addition to proportional truth-conditions, in these trials, the increase in overall rating in this condition is explained.

3.3. Experiment 3: Self-Paced Counting

Experiment 3 aims to investigate the verification process of *most*-statements with greater precision than Experiments 1–2. In addition to gathering information about Yes/No responses to trials containing *most* and *more than half*, Experiment 3 targets the verification process itself by looking at reaction time data. To gather fine-grained information about the verification process, we adopt the Self-Paced Counting (SPC) methodology of Hackl (2009). We introduce target trials which contain either *most* or *more than half* and dot arrays that contain either two colors or three colors, using the same target items as in Experiment 1 when possible.

Solving a verification problem that involves counting the number of objects in a scene is a complex task with many degrees of freedom, especially when the scene is displayed all at once. This makes it rather difficult to relate an observed difference in verification to a difference in linguistic form. To sidestep this difficulty, SPC reveals a given scene incrementally in a step-by-step and self-paced fashion similar to the widely used Self-Paced Reading methodology.

In a typical trial, subjects read a sentence such as (14) or (15) displayed on a computer screen. The sentence is paired with an array of up to 18 figures (circles, squares, triangles, stars) in different colors (blue, red, green, yellow, white) which are initially covered by large hexagons (screen 1 in Figure 7).

(14) Are most of the dots blue?

(15) Are more than half of the dots blue?

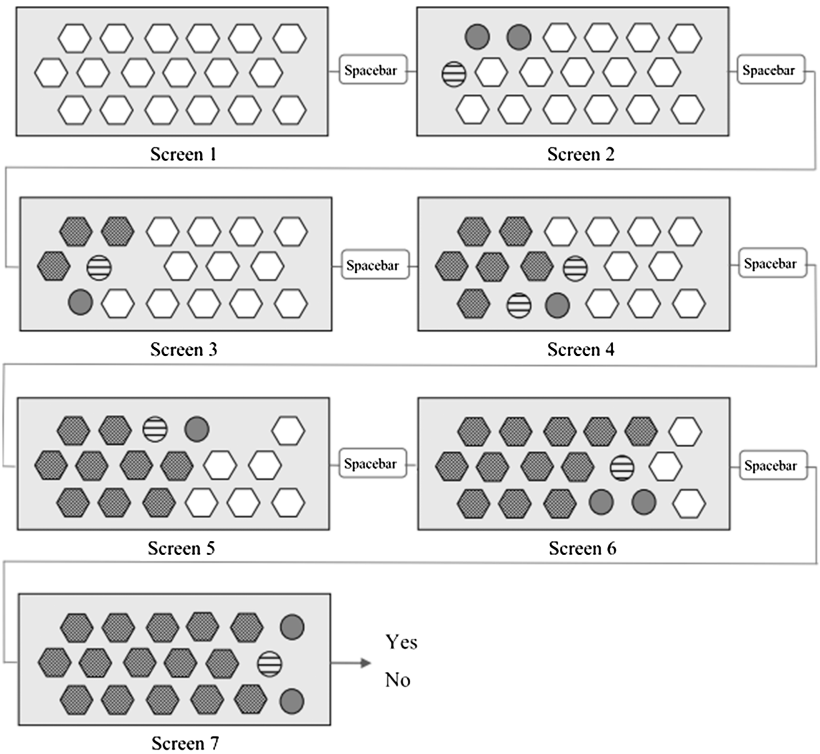


Figure 7: The sequence of events in a Self-Paced Counting trial.

As subjects press the spacebar the figures are uncovered in increments of two or three, while previously seen figures are masked. The opening sequence of the arrays varies but arrays are always uncovered from left to right. Once subjects gather enough information to answer the question asked about the array, they can respond "Yes" or "No" by pressing the appropriate response key on a keyboard, and they are encouraged to answer as reliably and as quickly as possible. The dependent measures are the reaction times on each screen and the proportion of correct answers.

3.3.1. Methods and Materials

Experiment 3 crosses two factors, Determiner and Color, with the levels "most" and "more than half," and "2-Color" and "3-Color," respectively. Determiner is treated as a between-subject factor to avoid one determiner contaminating the chosen verification strategy of the other. There were 24 *more than half/most* target items. 12 items were 2-Color and 12 were 3-Color. In each category, 6 items were true and 6 were false. The target trials in the "most" condition used the same set of dot arrays as target trials in the "more than half" condition; these conditions thus differed only in the statements paired with the arrays.

Each target item contained only one kind of figure out of circle, star, square, and triangle, and 2 or 3 colors out of five possibilities: blue, red, green, yellow, and white. The total number of figures in each item was between 14 and 18 figures, with 18 hexagonal covers and 6 spacebar presses required to uncover all of the figures in one item (7 screens in total, see Figure 7). The difference between true and false items was always of one figure (true ratios were 9:8, 8:7 and false ratios were 8:9, 7:8). Each "handful" of dots uncovered in a single screen contained at least one blue dot. In target trials that contained a third color, that color was introduced in screen 5 (i.e., after 4 spacebar presses) and there were always exactly 4 figures in the third color in screens 5–7. In all of the target trials, participants could not decide whether the blue dots outnumbered the non-blue dots in the array until after having uncovered the whole array.

There were 96 filler items in total, of which 60 were 2-Color items and 36 were 3-Color items. Half of the items were true and half were false. Each item contained only one kind of figure and colors were varied across items. There were 24 *more than n* filler items which were constructed in the same way as the target items and contained the same ratios. Half were 2-Color and half were 3-Color. These items are used as controls in order to verify that subjects understood the SPC task and were successful in verifying items with similar designs as the target *most* and *more than half* trials.

In the rest of the filler items, the screen in which the third color appeared was not controlled, so that all the colors were introduced randomly, and the number of figures in the third color was also varied. The screen at which a decision could be made was varied. In some cases it was possible to answer before uncovering the whole array while in others sufficient knowledge for answering the questions was available only after the last set of figures had been uncovered. Filler arrays ranged from 13 to 18 figures and were paired with questions given in (16)–(19).

(16) Are {exactly/about/more than} {3/4/5/6/7/8/9/10/12} of the circles blue?

(17) Are {about/more than} $\{\frac{1}{3}/\frac{2}{3}/\frac{1}{4}/\frac{3}{4}/\frac{1}{5}\}$ of the circles blue?

(18) Are {about/more than} {10/20/50/75/90} percent of the circles blue?

(19) Are {many/few} of the circles blue?

The items in (16) represent simple counting determiners. Each item in this category is paired with a question containing one determiner and one number out of the choices presented in (16); the items in (17) represent proportional determiners with fractions; the items in (18) represent proportional determiners with percentages; and the items in (19) represent proportional determiners with *few* and *many*. There were 36 proportional filler items, 12 for each of the three categories in (17)–(19), where *few* and *many* were only paired with 2-Color trials and for each of the other two categories, half of the trials were 2-Color and half were 3-Color. Subjects were given 8 practice items with filler determiners at the beginning of the session. The stimuli were presented in a pseudorandom order using software developed especially for the presentation of SPC items, coded in MATLAB and using Psychtoolbox to present the stimuli, record the time between successive spacebar presses, the time it took to press an answer key, the screen at which the answer key was pressed, and whether the given answer was correct. Stimuli were presented on iMac computers at the Behavioral Research Lab at MIT to 54 English native speakers, who received \$10 in cash as compensation. Twenty-seven subjects were assigned the "more than half" condition and 27 subjects were assigned the "most" condition.

3.3.2 Rates of Yes/No Responses

3.3.2.1. Results. The SPC design of Experiment 3 yields two separate dimensions along which the effect of the superlative reading of *most* can be observed. We start by surveying the Yes/No response rates in the experiment. We take *more than half* as a baseline against which *most* can be compared: since it unambiguously only has a proportional reading, we can measure the difference between its behavior pattern and the behavior of *most*.

Five subjects who answered less than 80% of the filler items (determiners other than "most" and "more than half") correctly were excluded from the analysis; 3 of them saw the "most" condition and 2 saw the "more than half"

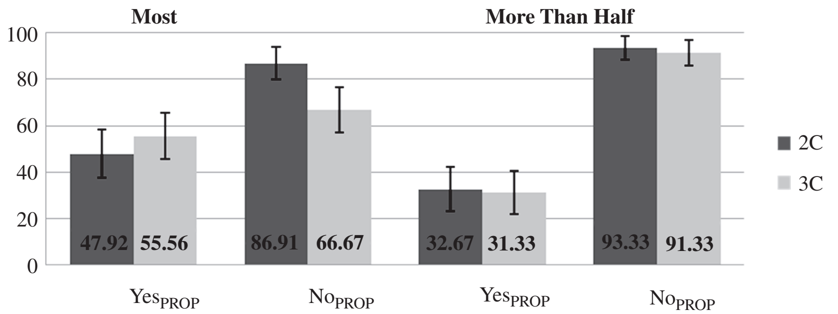


Figure 8: Average response rates for *most* and *more than half* in Experiment 3.

condition. Of the remaining 49 subjects, 24 subjects are in the "most" condition and 25 are in the "more than half" condition. The mean accuracy rate on filler items among the 24 "most" subjects is 91.36%, and the mean accuracy rate on filler items among the 25 "more than half" subjects is 90.75%. Figure 8 shows the (percentage) response rates to Yes_{PROP} and No_{PROP} trials for *most* and *more than half* broken down by number of colors in the arrays. The error bars represent standard errors.

For both *most* and *more than half*, we observe a very high response rate for No_{PROP} items but a surprisingly low response rate for Yes_{PROP} items. We also observe that the Color condition affects the verification of *most*, but it does not affect *more than half*.

To examine these trends statistically, we fit a mixed effects logit model with random subject and item effects to the answers for "most" and "more than half" separately. The model predicts the log-odds of the correct versus incorrect answers according to the proportional reading based on two categorical predictors, Correct Answer By Proportional Truth Conditions (Yes vs. No; CorrAnswer) and Color (2C vs. 3C).[12] The model includes random intercepts for both subjects and items, and a random slope of the effect of CorrAnswer for subjects.

The summary of the fixed effects of this model is given in Table 5 ($N = 576$; quasi-log-likelihood $= -282.8$). The correlations between fixed effects are all within ± 0.25. We observe the following results for *most*: A main effect of Truth, such that participants answered "No" to No_{PROP} items with significantly higher rates than they answered "Yes" to Yes_{PROP} items. A main

[12]The R code is given below and the predictor variables were coded as follows and centered:

(i)	CorrAnswers:	Yes = 0.5	No = −0.5
	Color:	3C = 0.5	2C = −0.5

(ii) lmer(Correct ~ CorrAnswer * Color+(1+CorrAnswer|SubjID)+(1|ItemID), distribution = "binomial") effect of Color, such that participants answered items in the 2-Color condition correctly ("No" for No$_{PROP}$ items and "Yes" for Yes$_{PROP}$ items) more often than items in the 3-Color condition and a Color × Truth interaction.

The random effect structure of the model was selected based on a series of step-wise model comparisons. A simpler model without the random effect of Correct Answer for subjects has a considerably lower quasi-log-likelihood −343.2, and a log-likelihood ratio test on the two models indicates a significant improvement brought about by the more complex random effect structure ($\chi^2(2) = 120.89$, $p < 0.001$). The inclusion of other random effects did not significantly ameliorate the performance of the model. The random effects are summarized in Table 6. The correlation between the effect of Correct Answer and intercept is −0.125.

A parallel mixed effects logit model was fit to the answers for *more than half*. We observe a main effect of Correct Answer such that participants responded with "No" to No$_{PROP}$ items at significantly higher rates than they responded with "Yes" to Yes$_{PROP}$ items. The effect of Color and the interaction are not significant. The fixed effects of this model are summarized in Table 7 ($N = 600$; quasi-log-likelihood = −214.8). The correlations between them are all within ±0.30.

The random effects are summarized in Table 8. The correlation between the two random effects for subjects is −0.23. Just as in the case of "most," a simpler model has a considerably lower quasi-log-likelihood, and a quasi-log-likelihood ratio test shows that the more complex random effect structure should be used ($\chi^2(2) = 86.346$, $p < 0.001$). Furthermore, more complex random effect structures did not result in a better model (the results of log-likelihood ratio tests are omitted here).

Table 5: Summary of the fixed effects for *most*.

Predictor	Coefficient	Standard error	Wald Z	p
Intercept	1.026	0.178	5.765	<0.001
CorrAnswer	−1.925	0.763	−2.521	<0.05
Color	−0.633	0.252	−2.516	<0.05
CorrAnswer × Color	2.226	0.504	4.413	<0.001

Table 6: Summary of the random effects for *most*.

Group	Random effect	s^2	Standard deviation
Item	Intercept	0.049	0.223
Subject	Intercept	0.311	0.558
	CorrAnswer	12.034	3.476

Table 7: Summary of the fixed effects for *more than half*.

Predictor	Coefficient	Standard error	Wald Z	p
Intercept	1.460	0.292	5.001	<0.001
CorrAnswer	−6.006	1.066	−5.635	<0.001
Color	−0.275	0.409	−0.674	0.501
CorrAnswer × Color	0.362	0.818	0.443	0.658

Table 8: Summary of the random effects for *more than half*.

Group	Random effect	s^2	Standard deviation
Item	Intercept	0.467	0.683
Subject	Intercept	0.559	0.748
	CorrAnswer	20.331	4.509

3.3.2.2. Discussion: Dominance and Charity. There are two trends in the response rates to target trials in Experiment 3. First, *most* and *more than half* react differently to the Color condition: only *most* is affected by the number of colors in the dot arrays, while *more than half* exhibits no sensitivity to color. This finding is in line with the findings of Experiments 1–2. Also in line with previous findings is the result that the proportional reading of *most* is more accessible than the superlative reading. This is revealed by a closer inspection of individual response patterns of the participants in Experiment 3 in the No_{PROP} condition: the observed behavior of *most* participants in Figure 8 indicates that a strategy compatible with superlative truth-conditions was chosen around 25% of the time, and a strategy compatible with proportional truth-conditions was chosen the rest of the time.

Second, the rates of "Yes" responses to Yes_{PROP} items for both determiners in Experiment 3 are surprisingly low, while the rates of "No" responses to No_{PROP} items are significantly higher. It appears that the participants in this experiment exhibit a No-bias that is not parallel to the findings of Experiments 1–2 and is not predicted by any theory of *most* discussed so far.

Although we propose to explain the sensitivity of *most* to the color manipulation in Experiments 1–3 by adopting a semantics for *most* which is inherently ambiguous between a proportional meaning and a superlative meaning, clearly the proportional meaning is preferred over the superlative meaning. In fact, to our knowledge, it has been the only reading observed in the literature for *most* in subject position. Only under special circumstances, such as the experimental design that made salient a partitioning of the dots to homogeneously colored subsets of dots, does the latent superlative meaning of *most* become available. Even then, it appears not to be available to all speakers all of the time.[13] We adopt the term Dominance for this state of affairs. The results of Experiments 1–3 teach us that some speakers choose to verify *most* statements according to the dominant reading, while others choose to verify according to the latent one. The ability to choose the dispreferred true reading of *most* over the strongly preferred false reading is explained under the Principle of Charity (Grice, 1975; Davidson, 1984; Hulsey, Hacquard, Fox, & Gualmini 2004, among many others):

(20) **The Principle of Charity**
When faced with an ambiguous sentence that is true on one reading and false on the other, language users prefer the reading of a sentence that makes it true.

If the Principle of Charity always prevailed in the competition between the two readings of *most*, we would expect any participant who can in principle access the superlative reading to answer "Yes" 100% of the time in the "superlative" conditions. This is, however, not the observed pattern of results in our experiments. Rather, it appears to be the case that in the competition between Dominance and Charity, one principle wins on some occasions and the other wins on other occasions for the very same speakers.

To explain the tension between Dominance and Charity, we adopt the argument in Musolino and Lidz (2003), according to which there is a competition between the bias to choose the interpretation that makes the sentence true (the Principle of Charity) and the difficulty in revising the initial parse of a sentence (Dominance). This is how Musolino and Lidz (2003, p. 288) put it: "one approach, inspired by recent models of sentence processing (Trueswell, Sekerina, Hill, & Logrip, 1999), would be to view relative processing difficulty and satisfaction of the principle of charity as probabilistic constraints exerting antagonistic forces: The greater the processing difficulty associated with a particular reading, the more likely it is that this difficulty will override the application of the principle of charity." That is, since in our case the strongly preferred reading of *most* is the proportional one, some speakers will tend to choose it despite being driven by Charity to prefer a true reading of the sentence. For other speakers, Charity will prevail over Dominance, making them switch to the dispreferred superlative reading of *most*, which makes the sentence true. Further, the competition between readings appears to happen for each *most* trial individually, instead of one time

[13]In Experiment 1, for example, we observe a significant preference for the covered box over the "superlative" picture. Further, 10 out of 23 subjects never chose the superlative picture, suggesting that the superlative reading is more difficult to access for those speakers, maybe even unavailable.

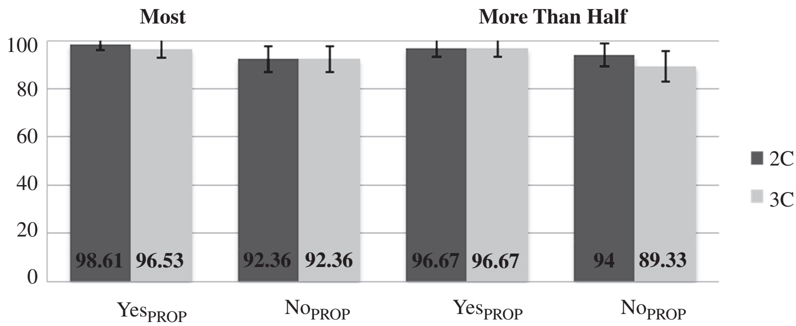


Figure 9: Response rates for *more than n* in Experiment 3 grouped by target determiner.

for the entire experiment, yielding different decisions for different trials within participants.[14]

3.3.2.3. Discussion: Confidence levels and the no-bias. One surprising result of Experiment 3 is the low accuracy rates exhibited by both *most* and *more than half* in Yes$_{PROP}$ trials, compared to the very high accuracy rates to No$_{PROP}$ trials, which can be described as a No-bias. To verify that the response rate results are not due to some general effect of the SPC task, we examine the response rates of the counting determiners *more than 7* and *more than 8*. The arrays in these conditions contained the same ratios of objects in the target and nontarget color as the *most* and *more than half* trials and hence happened to be true in the exact same cases as the proportional determiners. We observe that all of our subjects have very high response rates for *more than n*, regardless of whether they saw "most"-target trials or "more than half"-target trials. This suggests that the response rates observed for *most* and *more than half* reflect a genuine effect of the difficulty inherent in verifying the proportional determiners and not a general peculiarity of the speakers who participated in this study (Figure 9).

Since basic counting determiners like *more than n* do not exhibit low accuracy rates in any condition, the low accuracy in the target conditions cannot be due to the overall difficulty of the task in this experiment. The low accuracy rates also cannot be attributed to the fact that the ratios of blue-to-non-blue dots used in Experiment 3 were very low, so that the counts

[14]In Experiment 1, only 3 out of 13 superlative-*most* speakers used the superlative strategy all of the time. The other subjects used that strategy only some of the time, but not consistently. In Experiment 3, only 2 speakers always used superlative truth-conditions, another 10 used that strategy some of the time, and 12 only used proportional truth-conditions. See Section 3.3.3 for more discussion of the classification of speakers in Experiment 3 into two groups according to whether they verified target trials according to superlative truth-conditions, or only according to proportional truth-conditions.

of the target objects were either just above or just below the minimal counts for truth, causing general uncertainty among participants. If the reason for the low response rates was that the ratio of blue dots to non-blue dots was so low that participants were unsure whether the two magnitudes that they were comparing were similar or distinct, we would expect them to have similar error rates for both Yes_{PROP} and No_{PROP} items. However, we observe very high response rates for No_{PROP} items and low response rates only in Yes_{PROP} items.

We suspect that some of our participants understood the SPC task to require that the number of target dots in a given array must be significantly higher than the standard of comparison, for a statement containing them to be true. That is, *most of the dots are blue* would be judged as true if the number of blue dots in the array is significantly above the number of non-blue dots.[15] The "significantly greater than" relation ("$\gg$") is a relation which holds between two numbers n, m, if n/m is greater than c, for some value c specified by the context such that $c > 1$. The variable c allows the discourse context to play a role in determining the interpretation of such utterances, by allowing some degree of variation in what constitutes a "significant" difference between the two cardinalities to be compared. This relation is plausibly at work in the more "conventional," non-strictly truth-conditional use of *most* A B, which is licensed when the proportion of As that are Bs is around 85% (Ariel, 2004). The main prediction of this formulation of "significantly greater than" is that a speaker is more likely to judge a *most* utterance true as the ratio between the cardinalities to be compared increases, which is to say that this ratio is more likely to exceed the contextually determined threshold c. This prediction is also in line with the recent findings in Solt (2011), who compares the usage of *most* and *more than half* in the Corpus of Contemporary American English (COCA) and finds that *most* is used in cases where the proportion of As that are B is greater by some margin than 50%.

Furthermore, it is plausible that "$\gg$" can also explain the low response rate of the *more than half* Yes conditions. Participants would be less likely to judge true a sentence such as *more than half of the dots are blue* if the criterion they employed was that the number of blue dots be "significantly greater than" half the total number of dots, as in (21). Since in our items the number of target dots was always either one above or one below the number of nontarget dots, participants were likely to judge both *most* and *more than half* items as false when they were in fact true under a strict interpretation of the determiner. At this point we do not understand why participants would adopt this criterion: it is not consistent with the findings in Solt (2011),

[15]In fact, a number of participants indicated during the debriefing that this is the way they understood *most* or *more than half*, whichever determiner they happened to see (recall that Determiner was a between-subject factor).

who shows that *more than half* is predominantly used in situations in which the ratio between the blue dots and the non-blue dots is in the range of just over 50% to around 65%. It appears that under the verification task in a Self-Paced Counting paradigm, participants require more separation than the sharp boundary that is observed for production of *more than half* statements.

(21) $[\![\text{more than half}]\!](A)(B) = 1 \text{ iff} |A \cap B| \gg \frac{1}{2}|A|$

An anonymous reviewer suggests an alternative interpretation of the low accuracy rates for Yes_{PROP} trials which relies on an interaction between the use of "covers" and the design of target trials in Experiment 3: recall that in each SPC trial, participants are asked to answer a question about a dot array that contains dots in two or three colors. Initially all the dots in the array are covered by large hexagons; participants "lift" the hexagons to uncover the dots underneath them incrementally by pressing the spacebar. There are always 18 hexagons regardless of the number of dots in the array. That is, some hexagons cover blue dots, some cover red or yellow dots, and some cover "blanks," or more precisely — have no dot underneath them. Target trials contained ratios of 9:8 or 8:7 blue-to-non-blue dots; therefore, they always contained at least one "blank." If participants used a verification strategy that relied on counting the number of blue dots and using the given information about hexagons to calculate the ratio of blue dots in the array, then we expect them to always answer *are most of the dots blue?* and *are more than half of the dots blue?* with "No." To answer "Yes," arrays would have to contain at least 10 blue dots — which was never the case in target trials in Experiment 3.

Although we cannot rule out the possibility that some participants used this strategy, we find it an unlikely explanation of our overall results for two reasons. First, participants were explicitly briefed to attend to the target figures and not to the number of covering hexagons. Moreover, since target questions were about *dots* and not about *hexagons*, and since none of the trials ever called for the comparison of different kinds of objects in an array (e.g., squares to dots), the linguistic form also does not support attending to hexagons in the verification process.

More importantly for our purpose, if *most* and *more than half* were both verified using a strategy that only required participants to attend to the number of blue dots, it would be very surprising that the number of non-blue dots should have any effect on the verification process. Under this interpretation of the results, it is unexpected that participants would ever choose to verify *most* (but not *more than half*) using a strategy that involves attending to three sets instead of just one. Consequently, our results support the conclusion that *most* and *more than half* have different semantic representations, only one of which could be sensitive to the number of colors in the array.

3.3.3. Reaction Times

Next we turn our attention to reaction time data. All RTs in nonfinal screens that are above 5 seconds are discarded; these account for 1.5% of the total data, and 2.62% of the data in the target conditions. Of the remaining data points, those RTs that are more than 1.5 standard deviations away from the mean are also discarded, where the standard deviation and the mean are computed per subject, per determiner, and per screen. These include 1.12% of the data in the target conditions.

At first glance, it appears that there are no interesting effects within the reaction time data for *most* and *more than half*. Figure 10 presents average reaction times for *most* and *more than half* (in seconds) grouped by Determiner ("most" vs. "more than half"), Color (2C vs. 3C), and Screen (1–6, where screen 1 is shown after the first spacebar press, screen 2 after the second spacebar press, and so on). The error bars represent standard errors.

We fit linear mixed models to the logarithm of the reaction times log(RTs) using Determiner and Color as fixed effects predictor variables. The random effect structure we consider includes random intercepts for subjects, items, and screens, by subject and by screen random slopes for Determiner, and a by screen random slope for Color. The correlation between the random slope for Determiner and the intercept for screens is not included.[16] A series of model comparison indicate that more complex random effect structures are not motivated (the results of log-likelihood ratio tests omitted here). Although the

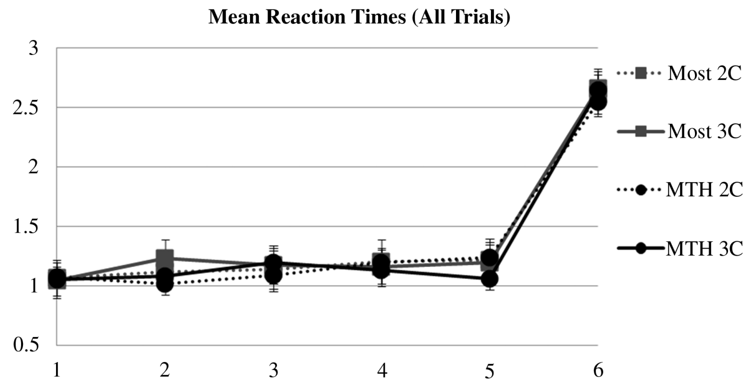


Figure 10: Average RTs for *most* and *more than half* in Experiment 3.

[16]The R code is given below and the predictor variables were coded as follows and centered.

(i) Determiner: Most = 0.5 More than half = −0.5
Color: 3C = 0.5 2C = −0.5

(ii) lmer(log(RT) ~ Color * Determiner+(1+Determiner | SubjID)+(1 | ItemID)+(1 | Screen) +(0+Color | Screen)+(0+Determiner | Screen))

Table 9: Summary of the fixed effects.

	Coefficient	Standard error	t
Intercept	0.070	0.175	0.398
Color	0.001	0.029	0.043
Determiner	−0.048	0.127	−0.377
Color × Determiner	−0.002	0.032	−0.074

Table 10: Summary of the random effects.

		Standard error	Standard deviation
Subject	Intercept	0.157	0.396
	Determiner	0.104	0.322
Item	Intercept	0.002	0.041
Frame	Intercept	0.162	0.402
	Color	0.003	0.058
	Determiner	0.002	0.046

correlation between the random intercept and the random effect of Determiner for screens is high (0.876), a log-likelihood ratio test with the model without this correlation indicates that its inclusion is justified ($\chi^2(1) = 8.949, p < 0.01$).

The fixed and random effects of this model are given in Tables 9 and 10 ($N = 6{,}576$, log likelihood $= -3{,}997$). The correlations among the fixed effects are within ± 0.204.

A log-likelihood ratio test on the model above and the model without the fixed interaction shows that the interaction is not significant ($\chi^2(1) = 0.006$, $p = 0.940$). Also the coefficients and the standard errors for the main fixed effects strongly suggest that their effects are nonsignificant.

Although the no-difference finding is surprising given the fact that the Yes/No response rates in Experiment 3 indicate that *most* and *more than half* behave differently in the experiment, with *most* being sensitive to the color manipulation and *more than half* being insensitive to the same manipulation, we note that this result may be caused by the fact that the data for *most* participants collapses two possible verification strategies, corresponding to the two different readings of *most*. Therefore, isolating "superlative" behavior from "proportional" behavior may allow us to observe that the two patterns are distinct from each other. To achieve this, we classify the participants in the *most* Condition into two populations, the *proportional* speakers and the *superlative* speakers, and compare their performance in the verification task. We use the post-hoc criterion in (22), which is based on the response patterns of participants in Experiment 1.

(22) If accuracy (2C-No$_{\text{PROP}}$) > accuracy (3C-No$_{\text{PROP}}$),
then superlative.
Otherwise: proportional.

Recall that we observed two patterns of behavior in that experiment: some participants always chose the covered box and never the "superlative" picture; other participants chose the covered box some of the time and the "superlative" picture some of the time. Overall, we can classify participants as "superlative speakers" if they verify *most* according to superlative truth-conditions at least some of the time; proportional speakers are defined as those speakers who don't verify *most* according to superlative truth-conditions. The criterion in (22) provides a classification of the participants in Experiment 3 by comparing the accuracy rates of the 2C- and 3C-No$_{\text{PROP}}$ conditions in this experiment. Speakers who use superlative truth-conditions for *most* at least some of the time will answer "Yes" more often to 3C trials than to 2C trials, since under the superlative reading the array is truthfully described by the *most* statement. By comparison, speakers who only use proportional truth-conditions will have the same accuracy pattern for the two Color conditions: trials in both conditions will be *false* for these speakers.

Note that there are two possible sources of "Yes" answers to trials in the 3C-No$_{\text{PROP}}$ condition. A speaker may verify a trial using proportional truth-conditions and answer "Yes" because of error, or she may verify a trial using superlative truth-conditions and answer "Yes" because that is the correct response according to these truth-conditions. Note further that because of the inherent complexity of 3C trials, we expect to find more errors in this condition compared to 2C trials. The criterion in (22) thus groups any general noise due to the color manipulation together with genuine superlative behavior. As a result of the new grouping, three speaker-populations are created in Experiment 3: 25 *more than half* participants, 12 proportional-*most* participants and 12 superlative-*most* participants.

Figure 11 shows the average reaction times (in seconds) for *most* and *more than half* in Experiment 3 grouped by speaker-population (superlative-*most* vs.

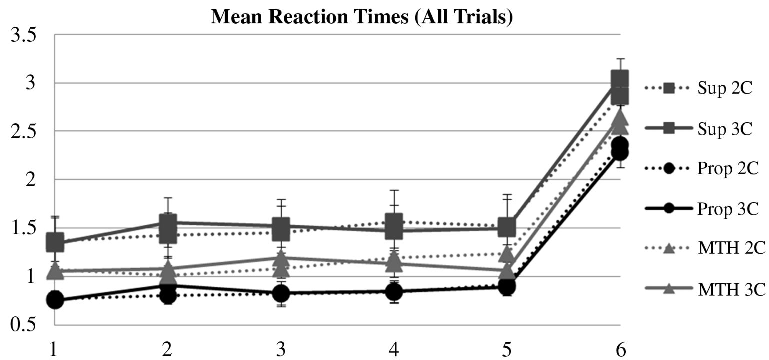


Figure 11: RTs for superlative-*most*, proportional-*most*, *more than half* in Experiment 3.

Table 11: Summary of the fixed effects.

Predictor	Coefficient	Standard error	t
Intercept	0.034	0.203	0.181
Color	−0.001	0.020	−0.025
SubjType	0.420	0.212	1.982
Color × SubjType	0.034	0.042	0.800

more than half vs. proportional-*most*), Number of Colors (2C vs. 3C) and Screen (1–6). The error bars represent standard errors. In this graph, we observe what appear to be three different patterns of behavior for the three speaker-populations isolated in our experiment. Below we will be interested in a statistical comparison of the two groups of *most* participants, who we suspect verified the same statements using different truth-conditions; and in a comparison of proportional-*most* participants and the *more than half* participants, who we suspect verified different statements according to equivalent truth-conditions.

3.3.3.1. Results: Proportional-*most* versus superlative-*most.* Just as above, we fit a linear mixed model to the logarithm of the reaction times log(RTs). We consider two predictor variables, Subject Type (Superlative vs. Proportional; SubjType) and Color.[17] The random effect structure we considered includes random intercepts for subjects, items, and screens, and by item random slopes for SubjType without the correlation with the random intercept. More complex random effect structures are not justified by log-likelihood ratio tests (omitted here).

The estimated coefficients for the fixed and random effects of the model are summarized in Tables 11 and 12 ($N = 3{,}183$, log-likelihood $= -1{,}896$). The correlations among the fixed effects are all within ± 0.024.

The estimated p-values via MCMC sampling ($n = 10{,}000$, conducted via the *pvals.fnc* function in the *languageR* package) indicates a significant main effect of SubjType, suggesting that the RTs of the two groups of the subjects are significantly different. The effect of Color and the interaction are not significant. The results of the MCMC sampling are summarized in Tables 13 and 14.

[17]Our predictors were coded and centered as indicated below and the R code is given in (ii).

(i) SubjType: Sup = 0.5 Prop = −0.5
Color: 3C = 0.5 2C = −0.5

(ii) lmer(log(RT) ~ SubjType * Color+(1 | SubjID)+(1 | ItemID)+(0+SubjType | ItemID)+(1 | Screen))

Table 12: Summary of the random effects.

Group	Random effect	s^2	Standard deviation
Subject	Intercept	0.267	0.517
Item	Intercept	0.001	0.033
	Determiner	0.005	0.073
Screen	Intercept	0.179	0.423

Table 13: Summary of the fixed effects of the MCMC sampling.

Fixed	MCMCmean	HPD95lower	HPD95upper	*pMCMC*	*p*
Intercept	−0.033	−0.283	0.337	0.813	0.856
Color	−0.001	−0.041	0.040	0.980	0.980
SubjType	0.418	−0.154	0.694	<0.01	<0.05
Color × SubjType	0.034	−0.048	0.118	0.410	0.424

Table 14: Summary of the random effects of the MCMC sampling.

Random	MCMCmedian	MCMCmean	HPD95lower	HPD95upper
Subject-Intercept	0.330	0.332	0.268	0.394
Item-Intercept	0.032	0.032	0.000	0.053
Item-SubjType	0.068	0.068	0.000	0.111
Frame-Intercept	0.324	0.337	0.193	0.518

3.3.3.2. Discussion. We find a significant main effect of speaker-population, such that proportional-*most* speakers verified target *most* trials significantly faster than superlative-*most* speakers. Unexpectedly, we do not observe an effect of the Color within the superlative-*most* population. To account for the lack of effect of the number colors within this group, we suggest that it may be caused by a combination of several reasons. First, recall that errors were grouped together with genuine "superlative" behavior in the superlative-*most* group; additionally, according to our grouping criterion, superlative-*most* speakers were classified as those who used superlative truth-conditions *some of the time*. Given that the proportional truth-conditions of *most* are dominant, we expect that these same speakers verified some *most* trials according to the proportional reading, and others according to the superlative reading. Consequently, the data for the superlative-*most* category represent a mix of two verification strategies, masking the effect of color which we expect to only be present in trials that were verified according to superlative truth-conditions.

As noted above, a second finding in the RT analysis is that proportional-*most* speakers verified target *most* trials significantly faster than superlative-*most* speakers. The reason that superlative-*most* speakers are slower, we suggest, is twofold. First, the superlative strategy involves making comparisons between homogeneously colored subsets of dots, instead of comparing the blue dots to the non-blue dots; in 3-Color trials this involves more comparisons, resulting in a slower verification process. A verification strategy that generalizes to the worst-case expects a similar slow-down for 2-Color trials. Second, we observe that the speakers we classify as superlative-*most* speakers use proportional truth-conditions in the verification of *most* some of the time, and superlative truth-conditions some of the time. This is the result of the conflict between dominance and charity discussed above. Recall that we concluded that it is not the case that superlative-*most* speakers choose just one verification strategy for all *most* trials. Rather, it appears that *most* is ambiguous between the two readings for these speakers, and that they must make a choice as to their verification strategy for each trial separately. Consequently, speakers for whom *most* is ambiguous are faced with two tasks in each of our target trials, compared to just one task for those speakers for whom *most* is unambiguously only proportional. They face a disambiguation task, in addition to the verification task.

To conclude, since speakers choose different truth-conditions for different trials, the graphs for superlative-*most* speakers represent a mix of strategies, making results difficult to interpret. Nonetheless, their behavior as a whole is clearly different from the behavior of proportional-*most* speakers, highlighting the existence of a second reading of *most* in addition to the proportional reading.

3.3.3.3. Results: Proportional-*most* versus *more than half*. We fit a parallel linear mixed model to the logarithm of the reaction times of the "more than half" trials and the "most" trials of the proportional participants using two predictor variables, Determiner (Proportional-Most vs. More than half) and Color.[18] The random effect structure we employ includes random intercepts for subjects, items, and screens and random slopes of Color and of Determiner for screens without correlations. More complex random effect structures are not justified by log-likelihood ratio tests, the results of which are omitted here.

[18]They are coded as indicated below and centered; the R code is given in (ii).

(i)	Determiner:	More than half = 0.5	Proportional-Most = −0.5
	Color:	3C = 0.5	2C = −0.5

(ii) lmer(log(RT) ~ Color*Determiner+(1|SubjID)+(1|ItemID)+(1|Screen)+(0+Color|Screen)+(0+Determiner|Screen)))

The fixed and random effects of this model are summarized in Tables 15 and 16 ($N = 5{,}057$, log-likelihood $= -3{,}041$). The correlations among the fixed effects are all within ± 0.120.

The estimated p-values via MCMC sampling ($n = 10{,}000$, conducted by the *pvals.fnc* function in the *languageR* package) indicate that Determiner has a significant effect. The negative coefficient of Determiner indicates that the proportional speakers were faster in verifying the sentence than the "more than half" participants. The results of the MCMC sampling are given in Tables 17 and 18.

Table 15: Summary of the fixed effects.

Predictor	Coefficient	Standard error	t
Intercept	0.008	0.171	0.049
Color	−0.004	0.034	−0.112
Determiner	−0.257	0.111	−2.320
Color × Determiner	−0.019	0.040	−0.474

Table 16: Summary of the random effects.

Group	Random effect	s^2	Standard deviation
Subjects	Intercept	0.092	0.304
Items	Intercept	0.003	0.052
Screens	Intercept	0.160	0.400
	Color	0.004	0.066
	Determiner	0.003	0.054

Table 17: Summary of the fixed effects of the MCMC sampling.

Fixed	MCMCmean	HPD95lower	HPD95upper	*pMCMC*	p
Intercept	0.007	−0.268	0.296	0.956	0.961
Color	−0.004	−0.094	0.089	0.921	0.911
Determiner	0.257	−0.454	−0.066	<0.05	<0.05
Color × Determiner	0.019	−0.096	0.062	0.624	0.635

Table 18: Summary of the random effects of the MCMC sampling.

Random	MCMCmedian	MCMCmean	HPD95lower	HPD95upper
Subject-Intercept	0.250	0.252	0.208	0.300
Item-Intercept	0.052	0.053	0.034	0.071
Frame-Intercept	0.298	0.316	0.171	0.505
Frame-Color	0.080	0.092	0.024	0.184
Frame-Determiner	0.065	0.074	0.012	0.162

3.3.3.4. Discussion. We find that proportional-*most* participants verified the target *most* trials significantly faster than the *more than half* participants verified the target *more than half* trials. As expected, the effect of color and the interaction are not significant. This result replicates the findings in Hackl (2009).

4. GENERAL DISCUSSION

In this section we discuss the results of Experiments 1–3 with regard to the debate over the correct analysis of *most*. We show how postulating a superlative reading of bare *most* in subject position explains the pattern of results observed in all three experiments and argue that under the lexical ambiguity view this reading can only be captured by proliferating lexical ambiguity. By contrast, under the structural ambiguity view the new reading requires only a minor adjustment to the proposal in Hackl (2009). The new ingredient is the claim that the comparison class argument of the superlative operator, C, whose content is by default identified with the extension of the plural NP of a bare *most* DP, can be further restricted under sufficient contextual support. We show that association with focus can be used as the grammatical means of providing this restriction and that doing so derives a restriction that results in a partitioned comparison class yielding superlative truth-conditions.

4.1. *Most* in Subject Position Has a Genuine Superlative Reading

Recall that the starting point for our experiments was the comparison between a lexical ambiguity view of *most* and a structural ambiguity view of *most*. We observed that the lexical ambiguity view predicts that bare *most* should only have proportional truth-conditions while the structural ambiguity view predicts that bare *most* should in principle be ambiguous between the two readings in (23) repeated from above.

(23) a. Most of the dots are blue
b. **Proportional truth-conditions**
$|\text{blue dots}| > |\text{non-blue dots}|$
c. **Superlative truth-conditions**
For each non-blue color Z, $|\text{blue dots}| > |\text{Z dots}|$

As noted above, for the superlative reading in (23)c to be detectable, dot arrays need to contain more than two colors and be structured so that the blue dots outnumber each set of non-blue dots but not their union. This is the general makeup of items in what we called the "superlative" condition: items in this condition contained dot arrays with ratios of, for example, 8 blue dots to 5 red dots and 4 yellow dots. If speakers can only verify a *most*-statement according to proportional truth-conditions, we expect them to judge *most of the dots are* *blue* as false in the “superlative” condition. If they can access superlative truth-conditions, they will be able to judge the same sentence as true in that condition.

All of our experiments used arrays in the “superlative” condition in order to probe for the existence of the superlative reading of *most* in subject position, using *more than half* as a baseline. The results are summarized in (24)–(26).

(24) **Experiment 1: Covered box**

Most:	The superlative picture was chosen 32.6% of the time. The covered box was chosen 67.4% of the time.
More than half:	The superlative picture was never chosen. The covered box was chosen 100% of the time.

(25) **Experiment 2: Rating study**

Most:	3C pictures in the “superlative” condition were rated significantly higher than the corresponding 2C pictures.
More than half:	No difference between the ratings of 3C trials and 2C trials.

(26) **Experiment 3: Self-Paced Counting**

Most:	Increase in “Yes”-responses to the “superlative” condition compared to the corresponding 2C baseline. Proportional-*most* speakers verified target trials significantly faster than superlative-*most* speakers.
More than half:	No change in “Yes”-responses to the “superlative” condition compared to the 2C baseline.

The results of all three experiments support the view that *most* in subject position can have a superlative reading. More specifically, we saw that *most* is sensitive to the color manipulation in all three experiments while *more than half* is not. This shows that the pattern of *most*-responses is not simply an indication of error behavior due to the overall difficulty of verifying proportional statements under our experimental settings. If this were the case, we would expect parallel error patterns for *most* and *more than half*. Moreover, we saw that the particular sensitivity to the color manipulation we observe for *most* tracks the expected behavior pattern of a verification strategy that is compatible with superlative truth-conditions. That is, we observe an increase in “Yes” responses in the “superlative” condition, which is expected if trials of that sort can be judged using superlative truth-conditions but would remain a mysterious error if *most* had only a proportional reading.

Even though our results show that a superlative reading of *most* in subject position exists, they do not suggest that the reading is freely accessible. In fact, the rate at which the superlative reading seems to manifest itself in our experiments is in line with the intuitions of those native speakers who can, upon reflection, intuit the reading, that it is clearly less preferable than the proportional reading. In Experiment 1 the covered box is chosen at significantly higher rates than the "superlative" picture, in Experiment 2 the "superlative" picture was rated significantly higher for *most* though unexpectedly low if the proportional and the superlative reading of *most* were equally accessible, and in Experiment 3 responses that are indicative of a superlative reading were given only about 25% of the time. Thus, across all three experiments proportional interpretations of *most* were prevalent while superlative interpretations were rare. This asymmetry could be due to dialectal variation (only some English speakers can access a superlative reading) or because the superlative reading, even though grammatical, is generally quite difficult to access. Our data do not allow us to distinguish these two possibilities. Whichever way turns out to be the case, the fact remains that our evidence points toward an unbalanced ambiguity of *most* in subject position, with the superlative reading being less readily accessible than the proportional reading.

The fact that bare *most* seems biased toward a proportional reading makes the latent superlative reading difficult to detect. We think that our experiments succeeded where probing native speaker intuitions about the meaning of *most* failed because the conditions under which our tasks were performed shifted the imbalance between the two readings somewhat. Our experiments all used some type of verification task (sentence-picture matching, sentence-picture rating, and SPC) and all of them exploited participants' general preference for "Yes" answers (agreeing with the experimenter) over "No" answers (disagreeing) when both are possible responses (Charity). Additionally, in all three experiments verifying according to superlative truth-conditions was feasible and in at least the first two even encouraged. In Experiment 1 participants were asked to choose between an overt "superlative" picture and a covert "proportional" picture in the critical trials; verifying according to superlative truth-conditions was encouraged simply because choosing the covered box is generally dispreferred. In Experiment 2 participants were asked to provide gradient sentence-picture matching judgments in the form of confidence ratings. This gave participants space to indicate that a sentence doesn't fit a picture very well yet better than a clearly false sentence. This is the contrast between the ratings of *most* and *more than half* for pictures in the superlative condition where the *more than half* sentence is clearly false and consequently given a very low rating while the *most* sentence is ambiguous and true under the latent superlative reading and thus given a higher rating. In Experiment 3 we used SPC, a method that has previously been shown to make verifying *most* sentences easier than verifying equivalent *more than half* sentences.

Of course, using verification tasks does not guarantee that the superlative reading can be brought to light. An example of a study that didn't find any evidence for the signature property of the superlative reading (sensitivity to the number of colors used across arrays) even though it used a verification paradigm with a similar color manipulation is Lidz et al. (2011).

4.2. Lidz et al. (2011)

The color manipulation used in Experiments 1–3 was based on the design considerations of an experiment conducted by Lidz et al. (2011). Their study examines the verification of the statement *most of the dots are blue* with regard to dot arrays, which were, unlike in our study, presented only for extremely short times (150 ms). The complexity of the dot arrays was varied systematically in the ratio of blue to non-blue dots (1:2, 2:3. 3:4, and 4:5) and in the number of colors used in the arrays (2,3,4, or 5). If determining the cardinalities of all the subsets of non-blue dots is part of the verification procedure of *most* statements, increasing the number of homogeneously colored subsets of non-blue dots should adversely affect the verification of *most*.

The results of this study show that subjects did not show increased difficulty in sentence verification as the complexity of the non-blue set increased. That is, participants were equally accurate in their responses across all four levels of the color manipulation. As a result, Lidz et al. (2011) conclude that the non-blue set could not have been directly involved in the verification procedure. They propose an analysis of *most* which is designed to show no effect of the complexity of the non-blue set, (27). Under this analysis, the cardinality of the complement set is calculated indirectly by subtracting the cardinality of the target set from the cardinality of the total set, rather than counted/estimated/measured directly.

(27) $[\![\text{most}]\!]\,(A)(B) = 1 \text{ iff} |A \cap B| > |A| - |A \cap B|$

For the purpose of our discussion, the proposal in (27) can be thought of as an instance of a lexical ambiguity view of *most* because bare *most* only has proportional truth-conditions and should therefore never be verified according to superlative truth-conditions. Further, under this denotation, *most* and *more than half* make use of the cardinalities of same sets in their computations: the set $A \cap B$ and the set A. It is therefore surprising, under this view, that the participants in our experiments showed sensitivity to the color manipulation with *most* but not with *more than half*.

Lidz et al. (2011) might very well be right that participants in their study verified *most* according to (27); clearly, a strategy consistent with the superlative truth-conditions of *most* was not used by the participants in this study. However, the fact that the strategy in (27) was used in this experiment does not entail that it is the only possible verification strategy of *most* statements or even that it is the default. It could be that a verification strategy that is consistent with superlative truth-conditions is in principle accessible to speakers, but that the experimental design in the Lidz et al. (2011) study made it a bad strategic choice to try to determine the size of all the subsets of non-blue dots in the arrays. In fact, previous work has independently shown that determining the size of the subsets of non-blue dots in arrays that contain 4 and 5 colors is very hard (Wolfe, 1998; Treisman & Gormican, 1988;

Treisman & Souther, 1985). Furthermore, Halberda, Sires, and Feigenson (2006) show that the set of all dots is always accessible in arrays with multiple colors. At least for the 4- and 5-color cases, then, subjects in the Lidz et al. (2011) experiment had to defer to a strategy like the one described in (27). Thus, this strategy had to be used at least 50% of the time. For the 2- and 3-color arrays, both a "superlative" strategy and a strategy as in (27) might be available. However, it seems plausible that subjects choose just one strategy for the entire experiment instead of switching between strategies according to the trial type. The strategy in (27) would thus be chosen, since it is guaranteed to be successful 100% of the time.

An anonymous reviewer asks if Lidz et al. (2011) could explain the results of Experiments 1–3 in this chapter using the mirror-image of the argument made above. That is, could one argue that (27) is the actual semantic representation of *most*, and that the results which show a sensitivity of *most* to a superlative reading are simply due to the design manipulations in our experiments which, in general, make verifying according to superlative truth-conditions more feasible? In a sense, this is exactly what we have suggested above except that in our proposal it is tied specifically to *most*. Indeed, one cannot say that our experiments encouraged superlative verification across the board since *more than half* showed no comparable sensitivity to the number of colors in our experiments. We need to explain, then, why our design manipulations affected *most* but not *more than half*. Under the proposal in (27) this is rather difficult because *most* and *more than half* unambiguously express proportional truth-conditions. Moreover, verification that is based on (27) would use the same elements as verification that is based on the natural semantics of *more than half*, namely the cardinalities of the sets A and $A \cap B$.

It seems, then, that the only move that a proposal based on (27) can make to account for our results is to assume that (27) isn't the only analysis of bare *most*. There has to be another meaning of *most* — presumably a superlative one given the pattern of results — that speakers can access at least some of the time. Since these two analyses are unrelated, such a proposal amounts to a lexical ambiguity view adding yet another layer of lexical ambiguity since it is no longer sufficient to stipulate that *the most* is a superlative and *most* is a proportional determiner. Such a view would also have to provide ad hoc conditions of use which correctly predict when each instance of *most* can be used. By contrast, on the structural ambiguity view of *most*, under which *most* is uniformly analyzed as a superlative construction, the new reading can be captured by a small adjustment of the comparison class argument of the superlative operator. Below we show that association with focus can be used as the grammatical means of executing this adjustment and that doing so derives a partitioned comparison class, which results in the superlative reading we have discovered. Since this adjustment targets the same parameter that governs the choice between the proportional and superlative readings of *most* in general, such a view offers a more principled explanation of our results.

4.3. Deriving the Superlative Reading of (Bare) *Most*

In this section we demonstrate how the superlative reading of *most* in subject position can be accommodated within the compositional semantics that treats *most* uniformly as a superlative construction. (28) and (29), repeated from above, summarize what a compositional analysis needs to derive for the basic cases, see Hackl (2009) for details.

(28) a. John talked to the most students.
b. $[\![$John talked to the most students$]\!] = 1$ iff $\exists d \exists X$[students(X) & John talked to X & $|X| \geq d$ & $\forall y \in C$[$y \neq$ John $\rightarrow \neg \exists Y$[students(Y) & y talked to Y & $|Y| \geq d$]]]
c. $C = D_e$

(29) a. John talked to most students.
b. $[\![$John talked to most students$]\!] = 1$ iff $\exists d \exists X$[students(X) & John talked to X & $|X| \geq d$ & $\forall Y \in C$ [students(Y) & $Y \perp X \rightarrow |Y| < d$]]
c. $C = {}^{*}[\![$students$]\!]$

(28)b is a standard treatment of a superlative meaning in terms of degree quantification: John is compared to other individuals in the comparison class, C, with regard to how many students he talked to — he needs to have talked to more students than anybody else in C for the sentence to be true. (29)b employs the same machinery but describes proportional truth-conditions because here, by assumption, the comparison class is identified with the extension of the plural noun *students,* indicated with the *-operator.[19] Thus student pluralities are compared to each other with regard to whether John talked to them or not and the sentence is true if there is a student plurality that John talked to which outnumbers all of its nonoverlapping alternatives.

The proposal, then, is that different choices for the comparison class result in either a regular superlative reading or a proportional determiner reading. To not lose track of the fact that (28)a and (29)a unambiguously express superlative and proportional truth-conditions respectively, the account needs to assume a mechanism that governs how C can be set. Hackl (2009) proposes that this mechanism is tied to the presence of the definite article. More specifically, following Heim (1985) and Szabolcsi (1986), Hackl (2009) postulates a movement-based mechanism according to which the superlative operator can be interpreted either inside the DP or outside. If *-est* is interpreted inside the DP, C is by default set to the extension of *students,* which results in proportional truth-conditions. If *-est* is interpreted outside of the DP, C is a set of alternatives to the subject (*John* in (29)).

[19]The distributivity operator "*" takes a predicate of atomic individuals P and returns that predicate which is true of a plural individual x if, for all of x's atomic subparts y, P(y) is true. Thus the extension of *P is the extension of P closed under sum formation (Link, 1983).

From this perspective the superlative reading of (30) is unexpected. The absence of the definite article should correlate with a setting of C that identifies it with the extension of the plural noun *dots* and this should result in proportional truth-conditions. However, it is easy to see that a simple adjustment to C can derive this reading. In particular, consider the resulting truth-conditions if C didn't contain all possible dot pluralities but only those that are homogenously colored. The value for C given in (30)c has this property, as it is the set of pluralities X such that all of X's atomic subparts are blue, or all of X's atomic subparts are red, etc.[20]

(30) a. Most of the dots are blue.
b. $[\![\text{Most of the dots are blue}]\!] = 1$ iff $\exists d \exists X[{}^*[\![\text{dot}]\!](X)\ \&\ {}^*[\![\text{blue}]\!](X)\ \&\ |X| \geq d\ \&\ \forall Y \in C\ [({}^*[\![\text{dot}]\!](Y)\ \&\ X \perp Y) \rightarrow |Y| < d]]$
c. $C = \{X: {}^*[\![\text{blue}]\!](X) \vee {}^*[\![\text{red}]\!](X) \vee {}^*[\![\text{yellow}]\!](X) \vee \ldots\}$

The net effect of the further restriction on C in (30)c is that the set of dot pluralities that count as relevant alternatives is a partition of the dots by color. Importantly, this is sufficient to predict truth-conditions that are no longer those of a proportional determiner. Instead, we predict a particular instance of a superlative reading, one that is based on a partitioning of the dots. According to this reading (30)a is true if there is a plurality of blue dots that outnumbers all homogenously colored alternatives. Relative to a picture as in Figure 12, for instance, the sentence should be judged true since there is no alternative plurality of dots which is as numerous as the blue dots.

On the other hand, if C is not partitioned by color and instead is simply the closure under sum formation, then we predict proportional truth-conditions and (30)a to be judged false relative to Figure 12, as less than half of the dots are blue.

The next question to ask is how this further restriction on C could come about. To address it, we sketch how the derivation of the superlative meaning

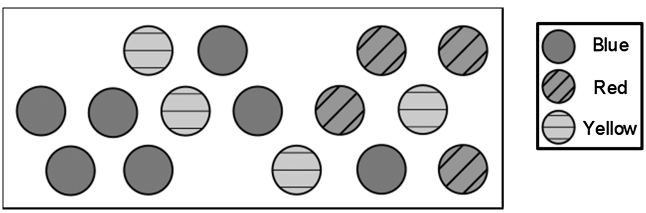


Figure 12: The "superlative" condition.

[20]Note that we assume two instances of the "*" operator, one each for *dot* and *blue*, which allows these predicates of atomic individuals to apply to pluralities.

of (30) can be modeled as resulting from the interaction of *-est* with focus. The sensitivity of superlative utterances to focus is discussed in Szabolsci (1986) and Heim (1999) (see also Farkas & Kiss, 2000, i.a.), and its effects can be detected in utterances such as (31):

(31) a. JOHN gave the most gifts to Mary.
 b. John gave the most gifts to MARY.

The meaning of (31) varies depending on whether sentential stress falls on the subject *John* or the dative argument *Mary.* Focusing the former implies that the individuals to be compared are those who gave Mary gifts, while focusing the latter implies that people to whom John gave gifts are to be compared. Thus, focus seems to play a role in determining the value of the comparison class argument of the superlative.

We suggest that the superlative reading of bare *most* in subject position is just another instance of this more general interplay between the superlative operator and focus.[21] Specifically, we suggest that focusing the color predicate *blue* in (30) can be used to restrict the comparison class which is by default set to the extension of **dot* resulting in a partition of the comparison class by color. To see how it could come about that focusing *blue* invokes the set of color predicates as alternatives, consider the derivation of (32), where we assume that the main predicate *blue* is focus-marked. Moreover, we assume that the salient alternatives to *blue* in the context of an array such as the one in Figure 12 are color predicates, and that they are generated in their pluralized form, (32)b. To get from this set of alternatives (a set of sets of pluralities) to a predicate that could restrict the comparison class (a set of pluralities) we apply grand union. As is easy to see, this produces the desired partitioning of the dots by color, (32)c.

(32) a. Most of the dots are BLUE.
 b. $\text{Focus\_Alt}(^*[\![\text{blue}]\!]) = \{^*[\![\text{blue}]\!], {}^*[\![\text{red}]\!], {}^*[\![\text{yellow}]\!], \ldots\}$
 c. $C = {}^*[\![\text{dots}]\!] \cap \cup(\{^*[\![\text{blue}]\!], {}^*[\![\text{red}]\!], {}^*[\![\text{yellow}]\!], \ldots\})$
 $= \{X: {}^*[\![\text{dots}]\!](X) \;\&\; ({}^*[\![\text{blue}]\!](X) \vee {}^*[\![\text{red}]\!](X) \vee {}^*[\![\text{yellow}]\!](X) \vee \ldots)\}$

Supplying this value for *-est*'s restrictor argument will derive superlative truth conditions, as discussed earlier in this section. Note that generating alternatives to *blue* before pluralization would result in a pathological comparison class. This is so because the grand union over a set of alternatives that range over atomic individuals only would result in a comparison class without any plural individuals in it. This would make any bare *most* statement trivially false

[21]We do not have the space here to present a fully explicit treatment of focus and the superlative operator's sensitivity to it. We hope that the sketch we do provide is clear enough to see how it might be executed, for example, within Rooth's (1992) framework of focus semantics. See Heim (1999) for discussion.

because if all individuals in C are atomic there cannot be an individual in C that is more numerous than all the others. Interestingly, generating the alternatives after pluralization seems to not only avoid this predicament, it also has the consequence that the resulting reading has the particular flavor of a partition-based superlative reading. This is obviously appropriate for sentences like (32). However, to see whether it is in general correct for superlative readings of bare *most* in subject position requires more research.

To summarize, we have sketched how the superlative reading of bare *most* in subject position can follow from the interplay of the underlying superlative semantics of *most*, distributivity, and mechanisms for the interpretation of focus. What the analysis sketched here leaves open, is why the latent reading is only sometimes detectable for certain speakers, and appears completely unavailable for others. From the perspective suggested here, we can restate this question and ask why imposing further restriction on the default setting of the comparison class using focus isn't freely available. We do not have a compelling answer to this question at the moment and need to leave it to future research.

5. SUMMARY AND CONCLUSION

We began this chapter by contrasting two competing views on the correct analysis of *most*: the lexical ambiguity view of *most* and the structural ambiguity view of *most*. While under the former view bare *most* is predicted to only have proportional truth-conditions equivalent to those of *more than half*, the latter view predicts that under certain circumstances *most* should give rise to a superlative interpretation. We tested this prediction using three different experiments involving sentence verification tasks. All three experiments support the conclusion that bare *most* in subject position is ambiguous between a dominant proportional reading and a latent superlative reading. To the best of our knowledge, this fact has not previously been noticed in the literature.

We use this observation to argue for two points, one methodological and one theoretical. The methodological point is that verification experiments can reveal aspects of the meaning of expressions that native speaker intuitions might miss. In particular, we suggest that the ambiguity of *most* in subject position has gone unnoticed because the two readings are not equally accessible: the dominant proportional reading masks the existence of the latent superlative reading. Consequently, probing for native speaker intuitions regarding entailment patterns associated with *most* may not provide the required context in which the ambiguity can be detected. Instead, specific task demands, such as those that obtained in our experiments, must be in place in order to systematically bring the latent superlative reading of *most* to the foreground.

The theoretical significance of our observation is that it provides compelling evidence for a decompositional analysis of *most* under which it is uniformly analyzed as a superlative construction. To account for the observation that bare *most* can give rise to superlative truth-conditions, the lexical ambiguity view would have to postulate yet another layer of lexical ambiguity: $most_{PROP}$, $most_{SUP}$, and *the most*, as well as ad hoc rules for when $most_{PROP}$ and $most_{SUP}$ should be used. The structural ambiguity view of *most*, on the other hand, can capture the ambiguity very naturally by further restricting the content of the comparison class argument of the superlative operator, C, whose content is by default identified with the extension of the plural NP of a bare *most* DP, yielding proportional truth-conditions. We develop an extension of the analysis in Hackl (2009) which suggests that the restriction on the content of C can be derived via the mechanism of association with focus, resulting in a partitioned comparison class that yields the desired superlative truth-conditions of *most* in subject position.

ACKNOWLEDGMENTS

Thanks to Akimitsu Hogge for programming the MATLAB script, Micha Breakstone for assistance with running Experiment 3, Michael Yoshitaka Erlewine for assistance with programming, Danny Fox and Greg Scontras for comments on the design of Experiment 3, and Jorie Koster-Hale for comments on Experiment 3. Special thanks to the audiences of CUNY 2011, CLS47, WCCFL29, SALT21, MIT LF reading group, Adam Albright, Irene Heim, Shigeru Miyagawa, David Pesetsky, Jeff Runner, and two anonymous reviewers for valuable suggestions and comments on the chapter. The research was supported by the *National Science Foundation* under Grant No. 0642748.

REFERENCES

Ariel, M. (2004). Most. *Language*, *80*, 658–706.

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, *4*, 159–219.

Bresnan, J. (1973). Syntax of the comparative clause construction in English. *Linguistic Inquiry*, *4,3*, 275–344.

Carpenter, P., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic model of sentence verification. *Psychological Review*, 45–73.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, 472–517.

Davidson, D. (1984). *Truth and interpretation*. Oxford: OUP.

Farkas, D., & Kiss, K. (2000). On the comparative and absolute readings of superlatives. *Natural Language and Linguistic Theory*, *18*, 417–455.

Grice, P. (1975). Logic and conversation. In: P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York, NY: Academic Press.

Hackl, M. (2009). On the grammar and processing of proportional quantifiers: *Most* versus *more than half*. *Natural Language Semantics*, *17*, 63–98.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, *17*, 572–576.

Heim, I. (1985). *Notes on comparatives and related matters*. Manuscript, University of Texas, Austin, TX. Semantics Archive. Retrieved from http://semanticsarchive.net/Archive/zc0ZjY0M

Heim, I. (1999). *Superlatives, MIT lecture notes*. Semantics Archive. Retrieved from http://semanticsarchive.net/Archive/TI1MTlhZ/Superlative.pdf

Huang, Y. T., Spelke, E., & Snedeker, J. *What exactly do numbers mean?* Manuscript.

Hulsey, S., Hacquard, V., Fox, D., & Gualmini, A. (2004). The question-answer requirement and scope assignment. In *Plato's problems: Papers on language acquisition*. MITWPL 48, 71–90. Cambridge: MIT Press.

Lidz, J., Pietroski, P., Hunter, T., & Halberda, J. (2011). Interface transparency and psychosemantics of *most*. *Natural Language Semantics, 19*, 227–256.

Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach. In: R. Bauerle, C. Schwarze & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (pp. 302–323). Berlin: de Gruyter.

Matthewson, L. (2001). Quantification and the nature of cross-linguistic variation. *Natural Language Semantics*, *9*, 145–189.

Musolino, J., & Lidz, J. (2003). The scope of isomorphism: Turning adults into children. *Language Acquisition*, *11*(4), 272–291.

Pearson, H., Khan, M., & Snedeker, J. (2010). Even more evidence for the emptiness of plurality: An Experimental investigation of plural interpretation as a species of implicature. *Proceedings of SALT*, *20*, 489–508.

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

Solt, S. (2011). *On measurement and quantification: The case of most and more than half*. Unpublished manuscript.

Szabolcsi, A. (1986). Comparative superlatives. In N. Fukui, T. Rapoport, & E. Sagey (Eds.), *Papers in theoretical linguistics* (pp. 245–265). MITWPL 8.

Szabolcsi, A. (2010). *Quantification*. Cambridge University Press.

Tanenhaus, M. K., Carroll, J. M., & Bever, T. G. (1976). Sentence-picture verification models as theories of sentence comprehension: A critique of carpenter and just. *Psychological Review*, *83*, 310–317.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*(1), 15–48.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: Human Perception & Performance*, *16*(3), 459–478.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134.

Wolfe, J. M. (1998). Visual search. In: H. Pashler (Ed.), *Attention* (pp. 13–73). Hove, UK: Psychology Press.

Yabushita, K. (1999). The unified semantics of mosts. *Proceedings of WCCFL*, *18*, 320–334.