

Up to n: Pragmatic inference about an optimal lower bound*

Ciyang Qing
Stanford University

Abstract This paper focuses on English directional modified numerals *up to n*, which triggers opposite inference patterns in *speaker-uncertainty* and *authoritative-permission* contexts. I propose that these opposite inference patterns are due to pragmatic inference about an unspecified semantic lower bound of *up to n*, based on its similarities to gradable adjectives and vague characteristics. The value of the semantic lower bound in different contexts is predicted by a general pragmatic principle of interaction between *informativity* and *applicability* independently motivated in previous probabilistic models on gradable adjectives.

Keywords: modified numerals, proximity inference, vagueness, optimal threshold model, probabilistic semantics and pragmatics, informativity-applicability trade-off

1 Introduction

This paper concerns the semantics and pragmatics of English *up to n* expressions, e.g., *up to 100*, which are called *directional modified numerals* in the literature Nouwen (2010). The main goal is to account for the contrast in the minimal pair (1).

- (1) a. You are about to meet up to 100 people.
b. You are allowed to meet up to 100 people.

On the one hand, (1a) is felicitous only in *speaker-uncertainty* contexts, where the speaker does not know the exact number of people that the listener is about to meet. A line of previous work has focused on how and why various types of modified numerals exhibit or lack this requirement (e.g., Geurts & Nouwen 2007; Buring 2008; Nouwen 2010; Coppock & Brochhagen 2013; Kennedy 2015). In addition, as Blok (2015a,b) observes, *up to n* in (1a) triggers a *proximity inference*, i.e., the speaker believes that the number of people is somewhere close to 100.

On the other hand, the most salient interpretation of (1b) is a permission to meet a number of people within the full range from 0 to 100, granted by a speaker who

* I thank Dan Lassiter, Cleo Condoravdi, Chris Potts for their valuable advice and support, Dominique Blok for kindly sending her SALT 25 manuscript and answering subsequent questions, the SALT 26 reviewers and participants for helpful comments and feedback, and my informants for the judgments.

has the authority.¹ I will refer to this as the *full-range inference* in *authoritative-permission* contexts.

We can see that in speaker-uncertainty and authoritative-permission contexts, *up to n* triggers opposite inference patterns: while the proximity inference in speaker-uncertainty contexts results in a narrow range of epistemic possibilities, the full-range inference in *authoritative-permission* contexts contributes to a wide range of deontic possibilities. A natural question arises: why and how does *up to n* trigger opposite inference patterns in these two contexts?

To answer this question, in the rest of the paper, I will first argue that *up to n* exhibits vagueness and highlight its similarities to vague gradable adjectives such as *tall*. Based on these similarities, I propose that *up to n* has a contextually determined semantic lower bound. Next I illustrate that the value of the semantic lower bound in different contexts is predicted by a general pragmatic principle of interaction between *informativity* and *applicability* independently motivated in previous work on gradable adjectives (Lassiter & Goodman 2013, 2015; Qing & Franke 2014a,b), and specify a probabilistic model that captures the opposite inference patterns.

2 Vagueness of *up to n*

In this section, I will apply Kennedy's (2007) characteristics of vagueness to argue that *up to n* exhibits vagueness, similar to vague gradable adjectives such as *tall*.

One main characteristic of vagueness is the existence of borderline cases and a lack of sharp boundaries exhibited by graded judgments (2).²

- (2) a. I expected John to be tall, but he is only/shorter than
5'5" / ... / 5'10" / (?)6' / ?6'2" / #7' (tall).
b. I expected to meet up to 100 people, but only/fewer than
10 / ... / 50 / (?)60 / ?75 / ??85 / #95 / #100 people are here.

The use of *but* in (2) indicates that the *but*-clause is not part of the speaker's prior expectation. Suppose John is a US adult male. Given that the speaker of (2a) had a prior expectation of John being tall, 5'5" is felicitous in the *but*-clause because a 5'5" US adult male is generally not considered tall, i.e., 5'5" is not in the interpretation of *tall* in (2a) and therefore not part of the speaker's prior expectation. In contrast, since 7' is certainly in the interpretation of *tall* in (2a), it is part of the speaker's prior expectation, and therefore it is infelicitous to use 7' in the *but*-clause. Meanwhile, for a borderline case such as 6', since it is unclear whether it is in the interpretation of *tall* in (2a), it is harder to decide whether 6' is felicitous in the *but*-clause.

1 When (1b) is used in a context in which the speaker is known to be uncertain about the exact number of people the listener is allowed to meet, it triggers a proximity inference similar to (1a).

2 Example (2b) is adapted from Blok (2015a) and generalized to test non-zero numbers in the *but*-clause.

We can see that *up to n* in (2b) patterns with *tall* in (2a): small numbers such as 10 are felicitous in the *but*-clause, suggesting that they are not part of the interpretation of *up to 100* in (2b). In contrast, large numbers close to 100, e.g., 95, are infelicitous in the *but*-clause, suggesting that they are part of the interpretation of *up to 100* in (2b). And finally, for numbers in between, e.g., 60, it is hard to judge whether they are felicitous in the *but*-clause, making them borderline cases for the interpretation of *up to 100* in (2b).

The choice of the borderline cases in (2b) is for illustrative purposes only. What is crucial is that the judgments are graded. The following naturally occurring examples in (3) further confirm that numbers small enough are felicitous in the *but*-clauses. Given that large numbers close enough to the upper bound are infelicitous in the *but*-clauses and we cannot identify sharp boundaries, there must be borderline cases somewhere in between, although they may well be different from person to person.

- (3) a. Vernell expected up to 10 vendors but only six materialized.
 b. It anticipated up to 40 cases would be mediated, but realised only 12.

Another characteristic of vague expressions is that they are susceptible to sorites paradoxes. For example, the premise that a 7' US adult male is tall seems plausible. So does the premise that a US adult male 0.5" shorter than a tall US adult male is still tall. However, with these two premises it follows that a 5'5" US adult male is also tall, which is implausible.

Now consider (4), a statement that former British Prime Minister David Cameron made to the House of Commons on refugees from Syria.

- (4) So Mr Speaker, we are proposing that Britain should resettle up to 20,000 Syrian refugees over the rest of this Parliament.

Given Cameron's proposal, it seems that the argument (5) has two plausible premises (5a) and (5b) and yet an implausible conclusion (5c).

- (5) a. Taking 20,000 refugees fulfills the proposal.
 b. $\forall i$, if taking i refugees fulfills the proposal, so does taking $(i - 1)$.
 c. # Therefore, taking 1 refugee fulfills the proposal.

Again, we have seen that *up to n* patterns with *tall* and exhibits vagueness. Large numbers close to the upper bound n are likely in its interpretation, very small numbers are likely not in its interpretation, and the judgments are graded: the larger the number, the more likely it is in the interpretation of *up to n*, but there is no clear-cut boundary.

Such characteristics suggest an analysis of *up to n* that makes use of its similarities to vague gradable adjectives. In the next section, I will incorporate this similarity into the semantics of *up to n*.

3 The semantics and pragmatics of *up to n*

In this section, I propose a semantics of *up to n* that incorporates its similarities to vague gradable adjectives, and argue that the opposite inference patterns of *up to n* is due to a pragmatic reasoning about the optimal lower bound.

3.1 An unspecified semantic lower bound

According to the degree-based semantics for gradable adjectives, the positive form of a gradable adjective *A* introduces an unspecified, contextually determined *standard of comparison* θ , such that *x is A* is true iff its degree of *A*-ness exceeds such a standard (Cresswell 1977; von Stechow 1984; Kennedy & McNally 2005; Kennedy 2007). For example, *John is tall* is true iff **height(John)** $\geq \theta$. Therefore, depending on θ , a degree *d* may or may not be part of the interpretation of the positive form. Assuming that contexts generally provide a probability distribution of θ , one can derive the probability of a degree *d* exceeding the standard θ to model the graded nature of vagueness (Lassiter 2011). For example, suppose that the contextual standard for being tall is uniformly distributed between 5'11" and 6'1". Small degrees such as 5'5" and 5'10" will always be excluded from the interpretation of *tall*, large degrees such as 7' will always be included, and an intermediate degree 6' is 50% likely to be included and is therefore a borderline case. In general, the larger the degree, the more likely it is in the interpretation of *tall*.

Based on the similarities between *up to n* and *tall*, I propose that *up to n* also has an unspecified semantic lower bound θ , whose value is contextually (probabilistically) determined. This straightforwardly captures the vagueness of *up to n* that we observe in the previous section.

3.2 Inquisitive semantics implementation

In addition to its similarities to gradable adjectives, *up to n* has other empirical properties, e.g., the requirement of speaker uncertainty in some linguistic environments (1a), which are different from gradable adjectives. I will adopt Coppock & Brochhagen's (2013) and Blok's (2015a) analyses of these properties, which are implemented in the inquisitive semantics framework (Ciardelli, Groenendijk & Roelofsen 2009, 2012), but note that my analysis of the opposite inference patterns of *up to n* does not hinge on the specific analyses of these other empirical properties or the formal framework used for their implementations.

Whereas a declarative sentence denotes a proposition (a set of possible worlds) in classical semantic theories, in inquisitive semantics, a declarative sentence denotes a set of propositions, i.e., a set of sets of possible worlds. A nice feature of inquisitive

semantics is that the classical denotation (truth condition) of a sentence can be retrieved by applying set union to its denotation in inquisitive semantics. This makes inquisitive semantics compatible with classical theories in terms of truth conditions, while having more fine-grained representations to capture the different inference patterns between expressions that have the same classical truth conditions.

I propose that *up to n* has the following semantics (6).

$$(6) \quad \llbracket \text{up to } n \rrbracket = \{ \lambda M_{\langle d,t \rangle}. \max(M) = k \mid k \in [\theta, n] \},$$

where θ is a contextual lower bound ($0 \leq \theta < n$).³

For example, the denotation of *up to 100* is in (7), which is a set of functions.

$$(7) \quad \begin{aligned} \llbracket \text{up to } 100 \rrbracket &= \{ \lambda M_{\langle d,t \rangle}. \max(M) = k \mid k \in [\theta, 100] \} \\ &= \{ \lambda M. \max(M) = \theta, \lambda M. \max(M) = \theta + 1, \dots, \lambda M. \max(M) = 100 \} \end{aligned}$$

θ is a contextual lower bound ($0 \leq \theta < n$)

For *you are about to meet up to 100 people* (1a), *up to 100* takes scope over the rest of the sentence, which is a degree property. After point-wise functional application, we obtain a set of propositions. The derivation is shown in (8).

$$(8) \quad \begin{aligned} &\llbracket \text{you are about to meet up to 100 people} \rrbracket \\ &= \llbracket \text{up to } 100 \rrbracket (\lambda d. \llbracket \text{you are about to meet } d\text{-many people} \rrbracket) \\ &= \{ p_\theta, p_{\theta+1}, \dots, p_{100} \} \end{aligned}$$

where θ is a contextual lower bound ($0 \leq \theta < n$) and p_i is the proposition that the listener is about to meet exactly i people.

The classical truth conditional content of (8) is the union of all the propositions, which is the proposition that the actual number of people who will attend the wedding, n_0 , is within the range $[\theta, 100]$. In other words, the informative content conveyed by (8) is that $n_0 \in [\theta, 100]$. The speaker asserts that the number is within this range. Meanwhile, since $\theta < 100$, the denotation of (8) always has at least two alternatives. According to Coppock & Brochhagen's (2013) Maxim of Interactive Sincerity, a cooperative speaker should not raise multiple alternatives if she already knows which one is true. Therefore the speaker of (8) will violate this maxim if she already knows the exact number of people, which explains the speaker-uncertainty requirement of (8).

On the other hand, for *you are allowed to meet up to 100 people* (1b) in authoritative-permission contexts, I adopt the common assumption in the literature that the permission modal scopes above *up to n* (Büring 2008; Coppock & Brochhagen 2013; Kennedy 2015). The derivation is shown in (9).

³ To improve readability, I omit the intensional types when they are irrelevant, e.g., the degree property $M_{\langle d,t \rangle}$ really is $M_{\langle s, \langle d,t \rangle \rangle}$. Also I will abbreviate $\max(\{d \mid M(d)\})$ to $\max(M)$.

Up to n: Optimal lower bound

- (9) $\llbracket \text{You are allowed to meet up to 100 people} \rrbracket = \{ \diamond \llbracket \text{you meet up to 100 people} \rrbracket \}$
 $= \{ \diamond \{ p_\theta, p_{\theta+1}, \dots, p_{100} \} \}$
where $0 \leq \theta < 100$ and p_i is the proposition that you meet exactly i people.

Furthermore, it is well known that permission modals scoping above a set of possibilities can trigger a *free-choice* inference (10), i.e., each possibility in the set is allowed. This is independently observed in studies of the interaction between permission modals and disjunctions (e.g., Kamp 1973, 1978; Zimmermann 2000; Kratzer & Shimoyama 2002).

- (10) $\diamond \{ p_\theta, p_{\theta+1}, \dots, p_{100} \} \rightsquigarrow \diamond p_\theta \wedge \diamond p_{\theta+1} \wedge \dots \wedge \diamond p_{100}$

However, various analyses of the free-choice inference disagree on whether the nature of this inference is semantic or pragmatic. I will not engage in the debate in this paper. The only assumption I will make is that the free-choice inference takes place before the pragmatic mechanism that I am going to propose next. If this inference is semantic entailment, then this is just a common assumption about the semantics/pragmatics interface. If this inference is itself pragmatic, it might seem unusual to assume that another pragmatic mechanism can happen after it. However, note that many pragmatic accounts of the free-choice inference assume that an expression can be used to defeat the implicature of an alternative expression, which effectively allows for the result of a pragmatic inference to feed into another pragmatic process anyways (e.g., Kratzer & Shimoyama 2002; Fox 2007; Franke 2011). Therefore it is viable to assume that the free-choice inference feeds into another pragmatic mechanism, even if the free-choice inference itself is pragmatic.

The proposed semantics of *up to n* allows us to explain the speaker-uncertainty requirement of (1a) by appealing to the Maxim of Interactive Sincerity. However, we still need to solve the main puzzle, i.e., explaining the opposite inference patterns in speaker-uncertainty and authoritative-permission contexts. Below, I will propose a pragmatic mechanism to contextually determine the unspecified lower bound θ .

3.3 Pragmatic reasoning about the unspecified lower bound

The pragmatic mechanism to contextually determine the unspecified lower bound I will propose is motivated by previous probabilistic models that aim at predicting the contextual standard of comparison for gradable adjectives (Lassiter & Goodman 2013, 2015; Qing & Franke 2014a,b). I will introduce the basic idea of those analyses below, and propose an adaptation to account for *up to n*. In the next section I will spell out a primitive probabilistic model that makes concrete quantitative predictions, which can be empirically tested in future studies.

Gradable adjectives are context-sensitive, e.g., a US adult male needs to be much taller to be considered tall than a five-year-old. This means that the standard of

comparison θ needs to be contextually determined and can change from context to context. In addition, *relative* gradable adjectives, such as *tall*, are vague (Kennedy & McNally 2005; Kennedy 2007).

As discussed in section 3.1, vagueness can be modeled probabilistically, i.e., instead of a fixed value, the context determines a probability distribution on the standard of comparison θ . Therefore, once the context is fixed, the goal is to explain and predict such a distribution, e.g., why the standard of comparison is most likely to be among certain degrees than others.

The crucial idea behind the probabilistic models mentioned above is that the standard of comparison should be a trade-off between *informativity* and *applicability*. For instance, when we talk about adult males in the US, the informativity of a standard of comparison θ is the information we will learn about the height of a person after hearing that he is tall, i.e., his height $h \geq \theta$. A higher θ makes the use of *tall* more informative, because it narrows down that person's possible heights more. However, at the same time a higher θ will make *tall* less applicable, in the sense that fewer people will be tall enough to qualify as tall. For example, if only people who are at least 7' can be described as tall, then it surely will be very informative when someone is described as tall, but unfortunately most of us will probably not have the chance to use it to describe anyone at all, which means that such a standard is not very applicable. More precisely, in this case applicability is the probability that the speaker can truthfully assert that a random person (sampled from the *contextual comparison class*, e.g., US adult males) is tall. The higher the probability, the more chances the speaker has to use *tall*, i.e., *tall* is more applicable.

From the perspective of efficient communication, expressions should ideally be both as informative and as applicable as possible. However, as discussed above, in the case of *tall*, the standard of comparison θ cannot optimize both at the same time. The consideration of informativity pushes the standard of comparison to larger degrees, while the consideration of applicability keeps it from being too large. As a result, the standard of comparison for *tall* is likely to be a degree that is relatively large in the context, but not unreasonably so. For example, for adult males in the US, 6'1" might be a good trade-off between informativity and applicability, so are 6' and 6'2", but not 5'5" (too uninformative) or 7' (too inapplicable). This is why the standard of comparison for *tall* in this context is likely to be 6'–6'2", but very unlikely to be 5'5" or 7'.

I propose that we can analyze the contextual lower bound of *up to n* by similarly consider the trade-off between informativity and applicability. However, the notions of informativity and applicability need to be adapted, depending on the type of contexts in which *up to n* is used.

First I consider *you are about to meet up to 100 people* (8). Let n_0 be the actual number of people that the listener is about to meet. The informative content is that

$n_0 \in [\theta, 100]$. A θ that is close to 100 will result in a narrower range, and therefore will make the sentence more informative.

Applicability is a little more complicated, because (8) needs to be both assertible and felicitous. We know that (8) is felicitous only when the speaker is uncertain about the exact number of people the listener will meet, n_0 . Since now the speaker does not know the value of n_0 , assertibility is not about the objective truth, but rather the speaker's subjective belief. Let G be the set of all numbers of guests that the speaker considers possible. In order for (8) to be assertible, we need $G \subseteq [\theta, n]$. Therefore, (8) is assertible iff $G \subseteq [\theta, n]$. Applicability is defined as the probability that (8) is assertible, i.e., the probability of $G \subseteq [\theta, n]$. In order to calculate such a probability, we need a probability distribution over G , which is from the contextual information about the speaker's level of uncertainty about the actual number n_0 . For example, if the speaker's information of the upcoming people is based on the number of invitations she has sent, most of which have not been confirmed, then she has a high level of uncertainty about n_0 and therefore G is likely to have many elements and the range $[\min(G), \max(G)]$ is likely to be wide. On the other hand, if the speaker just saw the crowd and spent some time to do some quick counting, then she has a low level of uncertainty and hence G is likely to have just a few elements and the range $[\min(G), \max(G)]$ is likely to be narrow.

Note that once the contextual distribution over G is fixed, a larger θ will always make the sentence less applicable. For instance, in the extreme case where $\theta = 99$, (8) asserts that the listener is about to see either exactly 99 or exactly 100 people, which is highly informative, but it is assertible only when $G \in \{99, 100\}$. Intuitively, it seems generally very unlikely that the speaker would know the number of people so well that they think this number can only be either 99 or 100. Therefore, the applicability, which is the probability that $G \in \{99, 100\}$, would be very low. Hence a θ that is too large will be dispreferred because it makes the sentence less applicable.

As a result, in speaker-uncertainty contexts, assuming that the speaker is overall informed but has residual uncertainty, when we take both informativity and applicability into account, the most likely lower bound θ for *up to n* should be relatively close to n (to be informative), but not unreasonably so (to still be applicable). Given that the lower bound θ is likely to be close to n , only degrees that are close to n are likely to be in the range $[\theta, n]$. Therefore *up to n* triggers a proximity inference in unembedded sentences.

Now I consider the authoritative permission contexts for *up to n* under permission modals, and take *you are allowed to meet up to 100 people* (9) as an example. Recall that I assume that the free-choice inference takes place before the current pragmatic considerations. Under this assumption, the informative content of (9) is the conjunction $\diamond p_\theta \wedge \diamond p_{\theta+1} \wedge \dots \wedge \diamond p_{100}$ (where p_i is the proposition that the listener meet exactly i people). Therefore, a smaller θ , which corresponds to more

conjuncts, will make the sentence more informative. For example, when $\theta = 10$, the listener will learn that meeting 10, 11, ..., 100 people are allowed, but when $\theta = 50$, the listener will only learn that meeting 50, 51, ..., 100 people are allowed, and remains uncertain about whether meeting 10, 11, ... or 49 people is allowed. We can see that the most informative θ would be 0, because it corresponds to the most conjuncts in the informative content.

In terms of applicability, note that (9) is assertible as long as the speaker has the authority, which is already the case for the class of relevant contexts we are considering. Therefore applicability does not favor any particular value of θ .⁴

As a result, consideration of informativity prefers a small θ , and applicability is not against it. Therefore the best θ would be 0 and the default interpretation is that the full range $[0, \theta]$ is allowed.

Therefore, I conclude that the opposite inference patterns of *up to n* in speaker-uncertainty and authoritative-permission contexts can be understood as the result of pragmatic reasoning about an unspecified semantic lower bound of *up to n*, subject to a general principle of trade-off between informativity and applicability. The two linguistic contexts differ in how the lower bound affects these two factors. This analysis has solved the main puzzle of this paper in section 1.

4 A probabilistic model

In this section I will propose a probabilistic model to show that the previous discussion can be formalized to make quantitative predictions. This primitive model has a lot of simplifying assumptions, but it suffices to illustrate the main concept.

For speaker-uncertainty contexts, the main challenge is to measure applicability quantitatively. This relies on the contextual information about the speaker's level of uncertainty about the actual number m . Ideally we would like to specify a distribution over the speaker's belief state G , i.e., the set of all the degrees that the speaker considers possible. However, in practice it is hard to specify such a big distribution, so I will make certain simplifications.

First, I will assume that the speaker's belief state G is a range $[a, b]$. The speaker chooses the upper bound n in her utterance of *up to n* based on a and b , and further chooses the implicit lower bound θ .

⁴ Note that this is making the simplifying assumption that authority is a binary concept, i.e., either the speaker has full authority or he does not. In situations in which there might be constraints from a higher authority, the speaker may only have *partial authority*. In such cases, the consideration of applicability will favor larger degrees, but the interaction with informativity will still result in a relatively small optimal lower bound. For example, given what we know about passwords, *you are allowed to use up to 20 characters for your password* might be permitting 6–20 characters. Similarly, since 0 is known to be ruled out in many contexts, *up to n* is often interpreted as permitting *1 to n*.

Up to n: Optimal lower bound

$$(11) \quad \begin{aligned} \text{a. } & p(m, a, b, \theta, n) = p(m, a, b) \cdot p(n \mid m, a, b) \cdot p(\theta \mid m, a, b, n) \\ \text{b. } & p(m, a, b, \theta, n) = p(m, a, b) \cdot p(n \mid a, b) \cdot p(\theta \mid a, b, n) \end{aligned}$$

According to the chain rule, we have (11a). Since the speaker does not know m and chooses the bounds based on her own belief a, b , the bounds θ, n are conditionally independent of m given a, b , and therefore (11a) can be reduced to (11b).

The model (11b) involves the choice of the upper bound n . Since we are more interested in the conditional probabilities $p(\theta \mid n)$ and $p(m \mid n)$, where n is already given, the choice of the upper bound n is not particularly relevant. Also, the main consideration in $p(n \mid a, b)$ is informativity, which is the same as in $p(\theta \mid a, b, n)$. Therefore, to simplify the model and highlight the informativity-applicability trade-off of θ , which is our main interest, I assume that the speaker always chooses $n = b$, i.e., n in *up to n* is the maximal number that the speaker considers possible. Intuitively this is very plausible: if 100 is maximal number of people that the speaker considers possible, then uttering *you are about to meet up to 120 people* is less informative than *you are about to meet up to 100 people* and there is no reason for a cooperative speaker to do that.⁵

This simplification helps us eliminate the variable b , and according to the chain rule, we obtain (12a).

$$(12) \quad \begin{aligned} \text{a. } & p(m, a, \theta \mid n) = p(m, a \mid n) \cdot p(\theta \mid m, a, n) = p(m, a \mid n) \cdot p(\theta \mid a, n) \\ \text{b. } & p(a, \theta \mid n) = p(a \mid n) \cdot p(\theta \mid a, n) \end{aligned}$$

We are interested in the conditional distribution $p(\theta \mid n)$, therefore we should marginalize over m and a . Marginalizing over m yields (12b). It has two parts: $p(a \mid n)$ encodes the contextual information about the speaker's level of uncertainty, and the second part reflects the informativity-applicability tradeoff.

First, recall that the sentence is assertible only when $[a, b] \subseteq [\theta, n]$, i.e., the semantic content of the utterance needs to be entailed by the speaker's belief state. This requires that $\theta \leq a$. Therefore, $p(\theta \mid a, n) = 0$ when $a < \theta$. When $a \geq \theta$, $p(\theta \mid a, n)$ depends on the informativity of θ , which is measured as the reduction in uncertainty (entropy) the lower bound θ contributes. Without the lower bound, the listener only knows that the number of people is between 0 and n , and therefore the entropy is $\log(n + 1)$, assuming a uniform prior. The lower bound θ narrows down the range to $[\theta, n]$, whose entropy is $\log(n + 1 - \theta)$. Therefore the reduction of entropy is $\log(n + 1) - \log(n + 1 - \theta)$, as shown in (13a). We can see that the larger the θ , the more informative it is.

⁵ Unless, of course, Quantity interacts with Manner. For instance, if 98 is the maximal number that the speaker considers possible, she might sacrifice a little bit of information and use the simpler form *up to 100*. This can be modeled by assigning higher costs to expressions with non-round numbers Kao, Wu, Bergen & Goodman (2014).

- (13) a. $\text{Informativity}(\theta) = \log(n+1) - \log(n+1-\theta)$
 b. $p(\theta | a, n) \propto \delta_{a \geq \theta} \cdot \text{Informativity}(\theta)^\lambda$
 c. $p(\theta | n) = \sum_a p(a | n) \cdot p(\theta | a, n)$
 $\propto \sum_a p(a | n) \cdot \delta_{a \geq \theta} \cdot \text{Informativity}(\theta)^\lambda$
 $= \sum_{a \geq \theta} p(a | n) \cdot \text{Informativity}(\theta)^\lambda$
 $= \text{Informativity}(\theta)^\lambda \cdot \sum_{a \geq \theta} p(a | n)$
 d. $\text{Applicability}(\theta) = \sum_{a \geq \theta} p(a | n)$
 e. $p(\theta | n) \propto \text{Informativity}(\theta)^\lambda \cdot \text{Applicability}(\theta)$

Now we can define $p(\theta | a, n)$ as in (13b), where $\delta_{a \geq \theta}$ is a delta function, which returns 1 if $a \geq \theta$ and 0 otherwise, and λ is a parameter that captures the importance of informativity in the choice of θ . If $\lambda = 0$, it means that informativity is not considered, and when $\lambda \rightarrow +\infty$, it means that informativity is the only consideration. Basically, (13b) says that the more informative a θ is, the more likely that it will get chosen, as long as the corresponding sentence is assertible, i.e., $\theta \leq a$.

Now we can marginalize over a , and plug in the definition in (13b) to simplify $p(\theta | n)$ in (13c). In the end we can see that $p(\theta | n)$ depends on the product of two terms. The first term is $\text{Informativity}(\theta)^\lambda$, which increases as θ increases. The second term is $\sum_{a \geq \theta} p(a | n)$, i.e., the probability of $a \geq \theta$, which is the probability that the corresponding *up to n* sentence is assertible. This is precisely the definition of applicability, as in (13d), and it decreases as θ increases. (13c) can be seen as the informativity-applicability tradeoff, as shown in (13e).

In order for the above model to produce actual quantitative predictions, we need to specify $p(a | n)$, i.e., the conditional probability of the minimal possible number considered by the speaker, given the maximal possible number n . This distribution is contextually determined, and depending on the context various assumptions can be made. Here, I will assume that this probability depends on n/a , i.e., the ratio between the maximal possible number, and the log ratio is normally distributed. This assumption seems plausible in many contexts, especially those that involve number perception, because previous work in psychophysics has shown that our perception is generally sensitive to ratios (see, e.g., Dehaene (2003) for more introduction and discussion) For example, our perception of the difference between 10 and 12 dots will be assumed more or less the same as the difference between 100 and 120 dots.

Assuming that $p(a | n) = \phi(\log(n/a))$, where $\phi(x)$ is a normal distribution shown in Figure 1(a), which corresponds to a contextual assumption that the max/min ratio is typically within 20%–50% and most likely around 35%.

When $\lambda = 4$, the predicted distribution of θ for *you are about to meet up to 100 people* is shown in Figure 1(b), but for a reasonable range of λ , the shape of the curve is qualitatively the same: a larger λ will shift the curve slightly to the right and

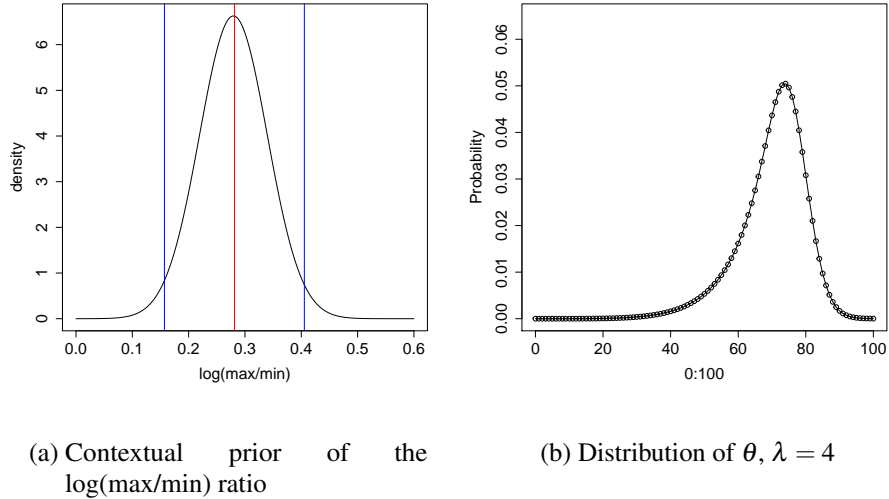


Figure 1 Model for *you are about to meet up to 100 people*

make it more concentrated, and a smaller λ will shift the curve slightly to the left and make it more flat. For instance, Figure 2(a) shows the prediction when $\lambda = 6$. This is the expected relation between the relative importance of informativity and the likelihood that a degree will be used as the lower bound for *up to n*. From Figure 2(a) we can see that the lower bound θ is seldom above 80 or below 50. This predicts that after hearing *you are about meet up to 100 people*, it would be inappropriate for the listener to say *no, there are only 80*, but it should be fine for the listener to say *no, there are only 50*.⁶

Note that the prediction is based on the ratio to the upper bound n rather than the difference. For example, we can see from Figure 2(b) that the shape of the curve is the same for *up to 10*. This result is also intuitively plausible.

We have seen how the probability of the lower bound θ can be inferred from the upper bound, i.e., $p(\theta | n)$. Quite often, we would also like to reason about the actual number m . However, when we first marginalize over θ in (12a) and then over a , we quickly realize that the lower bounds do not seem to have any impact on the actual number, as shown in (14).

$$(14) \quad \text{a. } p(m, a, \theta | n) = p(m, a | n) \cdot p(\theta | m, a, n) = p(m, a | n) \cdot p(\theta | a, n)$$

⁶ Keep in mind that this prediction is based on the specific contextual assumption about the speaker's level of uncertainty (around 35%, which is quite high). If the listener believes that the level uncertainty is lower, then *no, there are only 80* would be a possible response.

- b. $p(m, a | n) = p(m, a | n)$ Marginalize over θ
 c. $p(m | n) = p(m | n)$ Marginalize over a

This might look surprising at first, but it is actually plausible. Note that the actual number does not directly determine the lower bound or the upper bound of *up to n*. Its influence on the bounds, if any, is through the maximal and minimal numbers in the speaker's belief. Therefore, if no additional assumption is made to link the actual number to the speaker's belief, there is no way to infer "backward" and calculate $p(m | n)$.

To see this more clearly, note that by Bayes' rule we have (15).

$$(15) \quad p(m | n) \propto p(m) \cdot p(n | m)$$

Here $p(m)$ is the prior of the actual number, and $p(n | m)$ is the probability that the maximal number that the speaker considers possible is n when the actual number is m . This is the link from the reality to the speaker's belief that we need to specify. In the extreme case where n and m are totally independent. For instance, in a context where the speaker chooses the commonly known absolute contextual maximum as the upper bound, e.g., the full capacity of the room $n = 100$, then the listener will gain no information at all from the utterance *you are about to meet up to 100 people*. In other cases, the speaker's belief is formed by a noisy observation of the actual number m , e.g., when talking about a place the speaker just left. In such cases, the probability $p(n | m)$ may be reasonably assumed to depend on the ratio n/m and the log ratio is normally distributed. Of course, the mean and variance of this distribution will be smaller than those of the log max/min ratio. Sometimes it might be reasonable to assume that n/m and m/a are independently distributed with identical distributions, which means that the mean and variance of n/a is twice as much as those of n/m . In yet some other cases, it might actually be most natural to directly estimate $p(m | n)$. For example, if the maximal number that speaker considers possible is based on the number of people who have responded "yes" or "maybe" to the invitation, then the listener can use his general knowledge about the typical attendance rate to directly estimate $p(m | n)$.

This means that depending on additional contextual information about $p(m | n)$ or $p(n | m)$, the listener's inferred distribution of the actual number m after hearing *up to n* can vary, and it is possible to infer, e.g., that 45 is most likely after hearing *up to 50*.

Now I consider the authoritative-permission contexts of using *up to n* under permission modals, and use *you are allowed to meet up to n people* as an example. The sentence is assertible as long as the speaker is assumed to have authority. Therefore, so long as the resulting meaning is sensible, the sentence will always be applicable therefore I will assume that the applicability of θ is a constant (16a).

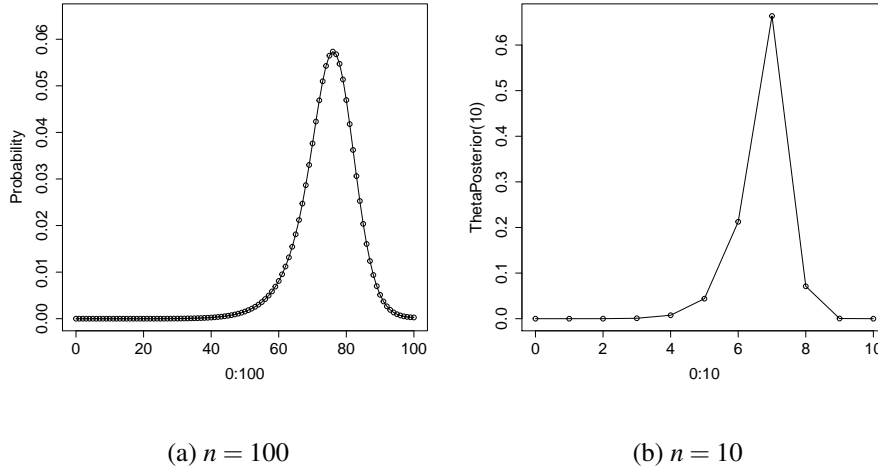


Figure 2 Distribution of θ for *you are about to meet up to n people*, with $\lambda = 6$

- (16) a. $\text{Applicability}(\theta) = 1$
- b. $\text{Informativity}(\theta) = (n + 1 - \theta) * \log 2 = n + 1 - \theta$
- c. $\text{Pr}(\theta) \propto \text{Informativity}(\theta)^\lambda \cdot \text{Applicability}(\theta)$

The informative content is the big conjunction that meeting exactly $\theta, \theta + 1, \dots, n$ people are all allowed. To simplify, I will assume that the listener initially is totally ignorant about whether meeting i people is allowed for any number of i , and they are all independent of each other. This means that for each i the entropy of the listener’s belief is $\log_2 2 = 1$. After hearing the sentence, the listener has no uncertainty about the status of $\theta, \theta + 1, \dots, n$ (since he learns that they are all allowed), which means the entropy of the listener’s belief about each of these numbers is now 0. Therefore the total reduction of entropy is $n + 1 - \theta$, as shown in (16b). Finally the trade-off between applicability and informativity is the same (16c), and I will use $\lambda = 6$.

Under the above assumptions, the predictions of the distribution of θ for *up to 100* and *up to 10* are shown in Figure 3.

In both cases, we can see that the most likely θ is 0 and the distributions are monotonically decreasing. This corresponds to the intuition that the most likely interpretation is that the full range is allowed, when there are no prior preferences introduced from the context.

I want to emphasize that the purpose of introducing the probabilistic model is mainly to illustrate that in principle we can make concrete, quantitative predictions about the contextual distribution of the lower bound θ of *up to n*. This will enable

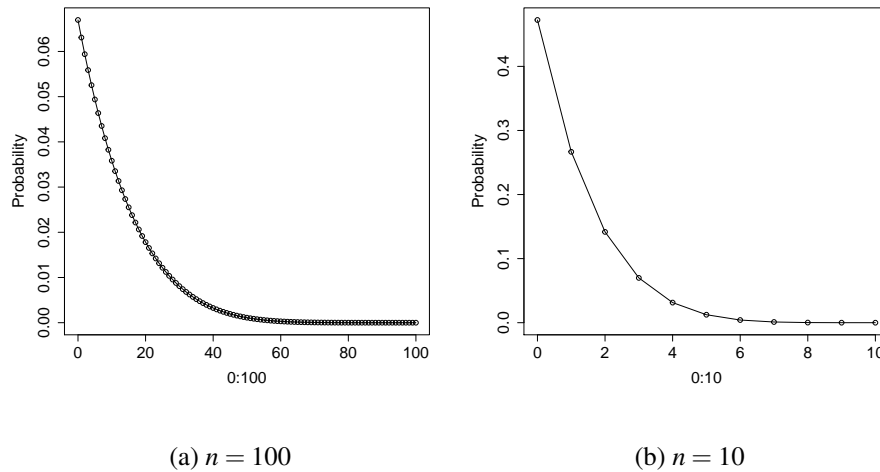


Figure 3 Distribution of θ for *you are allowed to meet up to n people*, with $\lambda = 6$

us to evaluate the extent to which the proposal that θ is determined by a trade-off between applicability and informativity captures the inference patterns of *up to n* in different linguistic environments. Of course, as we have seen, it is not a trivial task to formalize a probabilistic model. In building the model, I need to make a lot of simplifying assumptions that may well be empirically incorrect. Future work is needed to improve the model and test its quantitative predictions.

5 Discussion and comparison with previous work

5.1 Generalization to universal deontic modals

The proposed analysis of *up to n* can be generalized to explain its inference patterns under universal deontic modals.

Nouwen (2008) claims that *up to n* is “not so happy with strong modals” (17).

(17) ?? Jasper is required to invite up to 10 children to his party.

Note that Nouwen intends *up to 10* to take scope below the universal modal in (17), which corresponds to the authoritative reading (the wide-scope speaker-uncertain reading is perfectly fine here). While it is true that many native speakers tend to find (17) odd, many similar naturally-occurring examples can be found (18).

(18) a. The squad must contain up to 25 players and have no more than 17 players who do not fulfil the Home Grown Player criteria.

Up to n: Optimal lower bound

- b. These essays should present the authors' arguments in a clear and structured manner and should include up to 5 references. "In My View" contributions are limited to 1200 words.
- c. It was required that the answers link to at least one and up to 3 references from the reference module.
- d. [Q & A] What do I do if I receive a "warning" within the EMPLOYEE STATE/PROVINCE field? This field may not be left blank and must contain up to 2 characters of text.

These examples do not seem as weird and suggest that the oddity of (17) cannot be explained simply as some semantic violation. As another example, in Cameron's proposal, as discussed in section 2 and repeated below in (19), *up to n* has a clear upper bound and a vague non-zero lower bound that seems to be quite close to the upper bound 20,000.⁷ Therefore it seems that the acceptability of *up to n* under strong modals largely depends on the context.

- (19) So Mr Speaker, we are proposing that Britain should resettle up to 20,000 Syrian refugees over the rest of this Parliament.

This high contextual variability is expected under the current analysis. Given that the semantic content of *up to n* has an unspecified range $[\theta, n]$, the weirdness of (17) can be explained by considering informativity and applicability.

- (20) a. $\llbracket(17)\rrbracket = \{\Box\{p_\theta, p_{\theta+1}, \dots, p_{10}\}\}$
 where p_i is the proposition that Jasper invites exactly i children to his party.
- b. $\Box\{p_\theta, p_{\theta+1}, \dots, p_{10}\} \models \bigwedge_{i \notin [\theta, 10]} \neg \Diamond p_i$
- c. $\Box\{p_\theta, p_{\theta+1}, \dots, p_{10}\} \rightsquigarrow \Diamond p_\theta \wedge \Diamond p_{\theta+1} \wedge \Diamond p_{10}$

First, as usual, we assume that in the authoritative reading the deontic modal scopes above *up to n*. This time it is the necessity modal taking a set of alternative propositions as its argument (20a). The classical logical property of deontic necessity \Box dictates that anything outside of the informative content of its complement is not allowed. Therefore, we know that (20a) entails that any number outside the range $[\theta, 10]$ is not allowed (20b). Deontic necessity modals scoping above a set of alternatives can also give rise to the free-choice inference, e.g., *you are required to eat an apple or an orange* implies that eating an apple is allowed and eating an orange is allowed. The result of the free-choice inference of (20a) is that inviting any number within the range $[\theta, 10]$ is allowed (20c). Recall that I assume the result

⁷ Note that after the proposal is approved, we can use *it is required that Britain resettle up to 20,000 refugees* to report Britain's commitment, which means *up to n* indeed takes scope below the modal.

of the free-choice inference feeds into the pragmatic mechanism that determines θ , therefore (20b) and (20c) together completely settle whether inviting exactly i children is allowed, for all the numbers. Therefore the informativity will be the same for any θ and the choice of θ totally relies on applicability. As discussed earlier, similar to permission modals, (20a) is assertible iff the speaker is taken to be authoritative. Therefore, without further contextual information, the sentence will always be applicable. Therefore any θ should be equally good. However, this means that the informativity-applicability tradeoff does not prefer any θ . As a result, (17) would be very ambiguous: if any θ is equally good, it would be hard to know what exactly is allowed or required.

For naturally-occurring examples, the listener can resort to background world knowledge to reasonably infer the intention of the speaker. This is much harder for a decontextualized sentence such as (17): without enough background knowledge about the kind of party Jasper has, we could not tell whether inviting only a few children is allowed. I suggest that it is this great ambiguity that renders (17) odd.

5.2 Range of the lower bound

The current analysis of *up to n* is very similar to Blok's (2015a) account, which also posits a contextually determined lower bound (she uses s to denote it).

However, there are two crucial differences. First, in Blok's semantics the lower bound s of *up to n* can never be 0. Since Blok also requires that $s < n$, her semantics predicts that *up to one* is *never* felicitous when 1 is the smallest non-zero number in the underlying scale. This prediction agrees with Schwarz, Buccola & Hamilton's (2012) same descriptive generalization, which they refer to as the *bottom-of-the-scale effect* (BotS). However, there are naturally-occurring examples of *up to one* in various linguistic environments (21), where *up to one* arguably means 0 or 1, contra Schwarz et al. (2012) and Blok (2015a).

- (21) a. You are allowed to bring up to one guest.
 b. The committee will submit up to one application.
 c. Each panel should consist of a convener, up to four presenters, and up to one respondent.

These examples suggest that 0 can be part of the semantic content of *up to n*, which means that the lower bound can be 0. BotS, which seems to hold mostly for simple episodic sentences, requires a pragmatic explanation.

Second, Blok only introduces the unspecified lower bound s in the semantics, without providing a pragmatic mechanism of how it is contextually determined or using the value of the lower bound to explain the proximity inference. In this

Up to n: Optimal lower bound

sense, the current analysis is an extension of her account and makes more concrete quantitative predictions.

5.3 Upper bound and comparison with *at most n*

A significant part of Blok's proposal that I have not discussed is the differences between *up to n* and *at most n*. In terms of the lower bound, I agree with her that *up to n* differs from *at most n* in that only the former has an unspecified lower bound. In terms of the upper bound, Blok proposes that, unlike *at most n*, *up to n* in fact does not impose a semantic upper bound, which is not what I assume in this paper. However, this is not a crucial assumption. As a crude descriptive generalization, I think in most cases *up to n* patterns with bare numerals. This observation suggests that the upper bound interpretation might come from the numeral part rather than *up to*. All I need to assume is that it then feeds into the proposed pragmatic mechanism.

6 Conclusion

In this paper, I have examined the opposite inference patterns of *up to n* in speaker-uncertainty and authoritative-permission contexts. I argue that this puzzle can be solved by postulating an unspecified, contextually determined lower bound θ in the semantic content of *up to n*, together with a pragmatic mechanism to determine θ by considering the trade-off between applicability and informativity.

In future work, I plan to conduct experiments to gather data on how people interpret *up to n* (and other modified numerals) in order to test and improve the probabilistic model.

Blok (2015a) surveys the counterpart of *up to n* in a variety of languages and proposes cross-linguistic generalizations. Given that the English data considered in this paper already suggest the need to revise her analysis, it would be informative to check whether we can find corresponding examples in other languages. From some preliminary, informal discussion with native speakers of several languages, I have some evidence that some of the English data in this paper have counterparts in other languages (e.g., there are similar counterexamples of the bottom-of-the-scale effect in French, Italian, German, and Persian). The discussion has also revealed that the matter can be rather complicated in some languages (e.g., Greek) and therefore we need to be very careful when making cross-linguistic generalizations.

References

Blok, Dominique. 2015a. The semantics and pragmatics of directional numeral modifiers. In *Semantics and Linguistic Theory (SALT)*, vol. 25, LSA and CLC

- Publications.
- Blok, Dominique. 2015b. The semantics and pragmatics of directional numeral modifiers, <http://www.dominiqueblok.org/work>. Unpublished manuscript.
- Büring, Daniel. 2008. The least *at least* can do. In Charles B. Chang & Hannah J. Haynie (eds.), *West Coast Conference on Formal Linguistics (WCCFL) 26*, 114–120. Somerville, MA: Cascadilla Proceedings Project.
- Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2009. Attention! Might in inquisitive semantics. In Satoshi Ito Ed Cormany & David Lutz (eds.), *Semantics and Linguistic Theory SALT 19*, 91–108.
- Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2012. Inquisitive semantics. *NASSLLI 2012 lecture notes*.
- Coppock, Elizabeth & Thomas Brochhagen. 2013. Raising and resolving issues with scalar modifiers. *Semantics and Pragmatics* 6(3). 1–57. doi:10.3765/sp.6.3.
- Cresswell, M. J. 1977. The semantics of degree. In Barbara Partee (ed.), *Montague Grammar*, 261–292. New York: Academic Press.
- Dehaene, Stanislas. 2003. The neural basis of the Weber–Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences* 7(4). 145–147. doi:[http://dx.doi.org/10.1016/S1364-6613\(03\)00055-X](http://dx.doi.org/10.1016/S1364-6613(03)00055-X).
- Fox, Danny. 2007. Free choice disjunction and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*, 71–120. London: Palgrave Macmillan UK. doi:10.1057/9780230210752_4.
- Franke, Michael. 2011. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics* 4(1). 1–82. doi:10.3765/sp.4.1.
- Geurts, Bart & Rick Nouwen. 2007. ‘At Least’ et al.: The semantics of scalar modifiers. *Language* 83(3). 533–559. doi:10.1353/lan.2007.0115.
- Kamp, Hans. 1973. Free choice permission. *Aristotelian Society* 74. 57–74.
- Kamp, Hans. 1978. Semantics versus pragmatics. In Franz Guenther & Siegfried Josef Schmidt (eds.), *Formal Semantics and Pragmatics for Natural Languages*, 255–287. Dordrecht: Reidel.
- Kao, Justine T., Jean Wu, Leon Bergen & Noah D. Goodman. 2014. Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences* 111(33). 12002–12007.
- Kennedy, Christopher. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30. 1–45.
- Kennedy, Christopher. 2015. A "de-Fregean" semantics (and neo-gricean pragmatics) for modified and unmodified numerals. *Semantics and Pragmatics* 8(10). 1–44. doi:10.3765/sp.8.10.
- Kennedy, Christopher & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81(2). 345–381.

- Kratzer, Angelika & Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. In Yukio Otsu (ed.), *Third Tokyo Conference on Psycholinguistics*, 1–25. Tokyo, Japan: Hituzi Syobo Publishing Company.
- Lassiter, Daniel. 2011. Vagueness as probabilistic linguistic knowledge. In Rick Nouwen, Robert van Rooij, Uli Sauerland & Hans-Christian Schmitz (eds.), *Vagueness in Communication*, 127–150. Springer.
- Lassiter, Daniel & Noah D. Goodman. 2013. Context, scale structure, and statistics in the interpretation of positive-form adjectives. In Todd Snider (ed.), *Semantics and Linguistic Theory (SALT) 23*, 587–610. LSA and CLC Publications.
- Lassiter, Daniel & Noah D. Goodman. 2015. Adjectival vagueness in a Bayesian model of interpretation. *Synthese* 1–36. doi:10.1007/s11229-015-0786-1.
- Nouwen, Rick. 2008. Directionality in numeral quantifiers: the case of “up to”. In Tova Friedman & Satoshi Ito (eds.), *Semantics and Linguistic Theory (SALT)*, vol. 18, 569–582. LSA and CLC Publications.
- Nouwen, Rick. 2010. Two kinds of modified numerals. *Semantics and Pragmatics* 3(3). 1–41. doi:10.3765/sp.3.3.
- Qing, Ciyang & Michael Franke. 2014a. Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In Todd Snider, Sarah D’Antonio & Mia Wiegand (eds.), *Semantics and Linguistic Theory (SALT) 24*, 23–41. Ithaca, NY: CLC.
- Qing, Ciyang & Michael Franke. 2014b. Meaning and use of gradable adjectives: Formal modeling meets empirical data. In Paul Bello, Marcello Guarini, Marjorie McShane & Brian Scassellati (eds.), *Cognitive Science Society 36*, 1204–1209. Austin, TX: Cognitive Science Society.
- Schwarz, Bernhard, Brian Buccola & Michael Hamilton. 2012. Two types of class B numeral modifiers: A reply to nouwen 2010. *Semantics and Pragmatics* 5(1). 1–25. doi:10.3765/sp.5.1.
- von Stechow, Arnim. 1984. Comparing semantic theories of comparison. *Journal of Semantics* 3(1). 1–77.
- Zimmermann, Thomas Ede. 2000. Free choice disjunction and epistemic possibility. *Natural Language Semantics* 8. 255–290.

Ciyang Qing
450 Serra Mall, Building 460
Stanford, CA 94305
qciyang@stanford.edu