A Computational Model of Quantification in Natural Language

by

Avril Kenney

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Author ....
Department of Electrical Engineering and Computer Science
May 29, 2012

Certified by ....
Joshua B. Tenenbaum
Professor of Cognitive Science and Computation
Thesis Supervisor

Accepted by ....
Dennis M. Freeman
Chairman, Masters of Engineering Thesis Committee

A Computational Model of

Quantification in Natural Language

by

Avril Kenney

Submitted to the Department of Electrical Engineering and Computer Science
on May 29, 2012, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Natural languages have various ways of expressing quantification, such as the English words "some" and "all." Different such words exist in different languages, and the same word can communicate quite different quantities depending on the context. This thesis presents a computational framework for modeling quantificational meanings and their use in communication. The model can represent meanings that depend on absolute amounts (e.g., two) as well as relative amounts (e.g., half of the total) and context-dependent amounts. It can also represent meanings with presuppositions. Communication between a speaker and a listener is modeled as single exchanges in which both participants have noisy perception of the actual state of the world, the speaker tries to communicate some quantity to the listener by using some word chosen to be informative, and the listener tries to infer the quantity using the word and the assumption that the speaker was being informative. The usage patterns predicted by the model are qualitatively similar to how the words are actually used. The model also shows that the sets of words in real languages result in more efficient communication than randomly selected sets of words with comparable meanings.

Thesis Supervisor: Joshua B. Tenenbaum
Title: Professor of Cognitive Science and Computation

Acknowledgements

I would like to acknowledge the following people for their (direct or indirect) contributions to this thesis.

- Josh Tenenbaum, for being my thesis advisor and trusting me to know what I was doing.
- Steve Piantadosi, for thinking my ideas are good and making them better, and for lots of helpful feedback on drafts of this thesis.
- Noah Goodman, for good advice and discussions.
- Mike Frank, for showing me how to do experiments and that it's okay to try crazy ideas.
- Rebecca Saxe, for teaching me how to be a scientist.
- Ted Gibson, for being my undergrad advisor.
- Irene Heim, for making me excited about quantifiers and compositional semantics.
- My family and friends, for the hugs.

Contents

Chapter 1

Introduction

In natural languages, there are words that express different quantities or amounts – for example, English has words such as "some" and "all". However, across languages there is a large amount of variation in which quantificational meanings are encoded as words, and there are many such meanings for which there are not known to exist words in any languages. This thesis presents a computational model of quantificational words, on the basis that the purpose of words is to effectively communicate meaning, given a noisy perception of the world.

Quantificational words are a useful class of words to study because, unlike others such as nouns or adjectives, they do not refer to specific objects, actions, or properties in the real world; they are abstract linguistic constructs. The overall frequency of their use is high and is in many ways independent of non-linguistic circumstances: unlike a content word such as "red," which is likely to be used only in situations where red things are relevant, a quantifier such as "some" could easily be used in almost any situation. Furthermore, quantificational words are a "closed class" of words: each language has essentially an unchanging set of them, which makes them easier to enumerate and compare across languages.

1.1 Quantificational Words

In English, "quantifiers" are often equated with "determiners" (which are, roughly, words that act as specifiers for nouns), but there is not currently a consensus on what words actually are "quantifiers." This project will therefore sidestep the issue of these formal categorizations and deal with "quantificational words," which will be defined loosely as single words used to indicate "how much" or "how many." This excludes many other determiners, such as those that specify "which" (for example, "this" or "my").

Much of the past research on quantification has focused on formally characterizing the syntactic and semantic details of quantifiers in specific languages (e.g., Barwise & Cooper, 1981; Keenan & Stavi, 1986). Languages vary widely in terms of how they express quantification. In many languages, quantificational words do not occupy the syntactic position of determiner as they (arguably) do in English. They are sometimes expressed as adverbs or modifiers (similar to the English "dogs are always animals" as opposed to "all dogs are animals") (Matthewson, 1996; Bach, Jelinek, Kratzer, & Partee, 1995). Other languages use nouns with meanings like "part" or "small amount" (Everett, 2005). This thesis aims to account for the *meanings* of quantificational words, regardless of their grammatical roles.

1.2 Efficient Communication

Recent work has shown that language is in many ways optimized for efficient communication. Particularly relevant to the domain of quantification are other domains in which a finite set of words must be allocated to a large continuous space of meanings. Cross-linguistics studies of words for colors (Regier, Kay, & Khetarpal, 2007) and for spatial relationships (Khetarpal, Majid, & Regier, 2009) have found that the set of words in any language tends to maximize informativeness, given people's perception of similarity within the domain.

Perception of quantities is known to be inexact, and in particular, it follows Weber's Law, which states that the smallest perceivable change in a stimulus is proportional to the size of the stimulus itself (Weber, 1846). When judging exact number, people's estimates are centered around the correct number, and the standard deviation of the estimates is approximately a constant proportion of that number (Whalen, Gallistel, & Gelman, 1999; Dehaene, 1999). In other words, the smaller a number is, the more precisely people can estimate it. Therefore it would be expected that words for smaller quantities should also be more precise.

Some attempts have been made to map quantificational words directly to numerical values (e.g., Bass, Cascio, & O'Connor, 1974). Others have tried to characterize words as distributions over quantities, where each word has a range of amounts it could denote but some are more likely than others (e.g., Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986). However, this approach can only be taken so far, because some quantificational words are explicitly context-dependent. For almost any quantity there exists some situation in which it could be called "many," but in any particular situation there is a more limited set of quantities that "many" likely applies to. People's interpretations of such words rely heavily on world knowledge and expectations. When judging the proportions referred to by words like "many" or "few," people's estimates vary depending on their prior beliefs, judging "few" to be a smaller proportion when the proportion expected is smaller (Moxey & Sanford, 1993b) or when the total set is larger (Newstead, Pollard, & Riezebos, 1987). Judgments can even be influenced by the grouping of visible objects or the similarity of other objects present (Coventry, Cangelosi, Newstead, Bacon, & Rajapakse, 2005; Coventry, Cangelosi, Newstead, & Bugmann, 2010).

Furthermore, even with well-defined literal meanings, words can convey additional information through pragmatic inference. Unlike in a domain such as color, where the meanings are fairly absolute, a quantity word can convey more information than its literal meaning because certain inferences are licensed by the non-linguistic context. For example, in English the word "some" is often used with the implication "some but not all," based on the reasoning that if "all" were true then the speaker would have used that word instead. Such implicatures can be explained by assuming that speakers choose words informatively, based on how well the words specify the intended referent (Frank, Goodman, Lai, & Tenenbaum, 2009; Frank & Goodman, 2012). Goodman and Stuhlmüller (2012) show that listeners' inferences are sensitive to the speaker's knowledge state (if the speaker cannot see all of the objects, the "not all" implicature for "some" is cancelled), and they develop a computational model that predicts this behavior as rational communication.

In addition to the quantities that they are making statements about, some quantificational words also have *presuppositions*. A presupposition is something that is assumed to be true in order for a statement to be evaluated at all. For example, the word "the" presupposes that there exists a unique thing for it to refer to, and so a statement such as "the king of France is bald" is neither true nor false but simply nonsensical, because there is no king of France. It is usually infelicitous to use a word without a presupposition when there exists another word with a satisfied presupposition and otherwise the same meaning (such as using "a" in a situation where "the" could be used). Heim (1991) postulates a "maximize presupposition" maxim to explain this observation. Presuppositions are generally assumed to be common knowledge to all conversational participants, so it would seem that they provide no information. However, there are certainly some contexts in which a listener is not aware of (or has forgotten) whether a presupposition is satisfied, but upon hearing a statement that presupposes it, gains that information (Schlenker, to appear). It can therefore be more informative to use words with presuppositions than those without presuppositions.

1.3 Overview

The purpose of this thesis is to produce a computational model that can account for what sets of quantificational words exist in natural languages and how they are used. Specifically, the model accurately predicts what words are likely to be used in what contexts, and it shows that the sets of words in existing languages are efficient for communication.

This work differs from previous approaches in that it takes into account people's real-world perceptual uncertainty, so although it uses strict logical representations (a word either applies or does not apply in a given situation, there are no distributions contained in the meaning), the model's distribution for how likely a word is to be used for particular quantities in practice can be substantially nonuniform.

Chapter 2 describes a computational model for informative communication of noisily-perceived quantities. Chapter 3 presents results from the model and compares its behavior to real-world languages. Chapter 4 summarizes the thesis and discusses possible future work.

Chapter 2

Computational Model

The basis of the model is that in the world there are various quantities that need to be communicated, and there are words with which to communicate them. A speaker is presented with a situation in which there is some *total set* of objects, and some subset of the total set is the *reference set*, whose magnitude he needs to convey using a word. A listener perceives the same total set but does not know the reference set, so she needs to use the speaker's word to guess the size of the reference set.[1] Both the speaker and the listener have imperfect perception of the exact quantities.

2.1 Representation of Meanings

The meaning of a word has two parts, the *presupposition* and the *assertion*. The presupposition constrains the size of the total set, and the assertion constrains the size of the reference set. Each of these two parts is represented as an interval, with an upper and lower bound (which can be contained in the interval or not – i.e., it can be open or closed). The endpoints of an interval are not necessarily numbers, but are one of three types of points: absolute numbers (such as two), proportions relative to the total amount (such as one-half), or a special point that refers to some context-dependent "default" proportion of the total set size. For computational reasons, the upper bound for presuppositions is infinity, and the upper bound for assertions is 100% of the total set size. All quantities get rounded to the nearest integer. Table 2.1 shows examples of how the meanings of some English

[1]For disambiguation purposes, the speaker will be referred to with masculine pronouns and the listener will be referred to with feminine pronouns.

quantifiers are represented. This representation certainly ignores some of the subtleties of natural-language meanings – for example, it cannot represent the difference between "all" and "every" – but it is sufficient to capture many of the important features (also see Section 4.2.1 for a discussion of how the model can be extended).

Word	Presupposition		Assertion	
	Bottom	Top	Bottom	Top
the	≥ 1	≤ 1	≥ 1	≤ 1
some	≥ 0	$< \infty$	≥ 2	$\leq 100\%$
many	≥ 0	$< \infty$	$>$ default %	$\leq 100\%$
all	≥ 0	$< \infty$	$\geq 100\%$	$\leq 100\%$

Table 2.1: The model's representations of the meanings of some English quantifiers.

2.2 Communication Scenarios

2.2.1 World Setup

For each communication exchange, three parameters specify the context: the total set size, the reference set size, and the "default" proportion to be used by meanings that depend on it in context. In the results reported here, the total set size follows a uniform distribution from 1 to some specified maximum, and the reference set size follows a uniform distribution from 0 to the total set size. The default proportion is uniformly distributed from 0 to 1, but for computational simplicity it is rounded to the nearest 0.05 (and cannot be 0). The speaker and listener both know the distributions of these parameters.

There are two types of perceptual uncertainty about the exact quantities in a given situation. First, there is some probability that the total set is not completely visible. If it is not, the size of the visible portion is distributed uniformly from 0 to the total set size (including 0 but not including the total set size), and the size of the visible portion of the reference set is distributed uniformly from 0 to the actual reference set size (including 0 but not including the reference set size) with the restriction that the visible reference set size cannot be larger than the visible total set size. Second, there is uncertainty about the exact quantity of a set. The quantity a person perceives is sampled from a Gaussian distribution whose mean is the actual visible quantity and whose standard deviation is some constant multiple of the mean (the constant is a model parameter, called the *coefficient of variation*, that can be specified). Values for the perceived set sizes are sampled separately by the speaker and the listener. Each person knows whether he can see the whole set, but he does not know whether the other person can. The default proportion is known exactly by both participants.

Table 2.2 defines the variables that will be used for the derivations in the remaining sections. Figure 2-1 shows a diagram of the directed graphical models (Pearl, 1998); each variable is a node, and an arrow from a node x to another node y indicates that y's distribution is conditionally dependent on x. The variables' values are actually sampled from the full distribution, but the speaker and listener each have their own simpler model of the other's behavior.

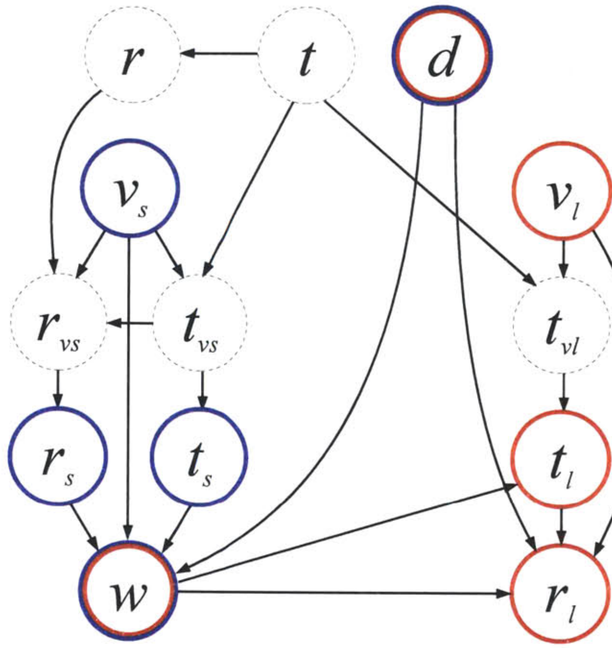
Variable	Definition
r	actual reference set size
r_{vs}	reference set size visible to speaker
r_s	speaker's perceived reference set size
r_ℓ	listener's guessed reference set size
t	actual total set size
t_{vs}	total set size visible to speaker
t_s	speaker's perceived total set size
$t_{v\ell}$	total set size visible to listener
t_ℓ	listener's perceived total set size
v_s	whether set is completely visible to speaker
v_ℓ	whether set is completely visible to listener
w	word
d	default proportion

Table 2.2: Variables used in the model.

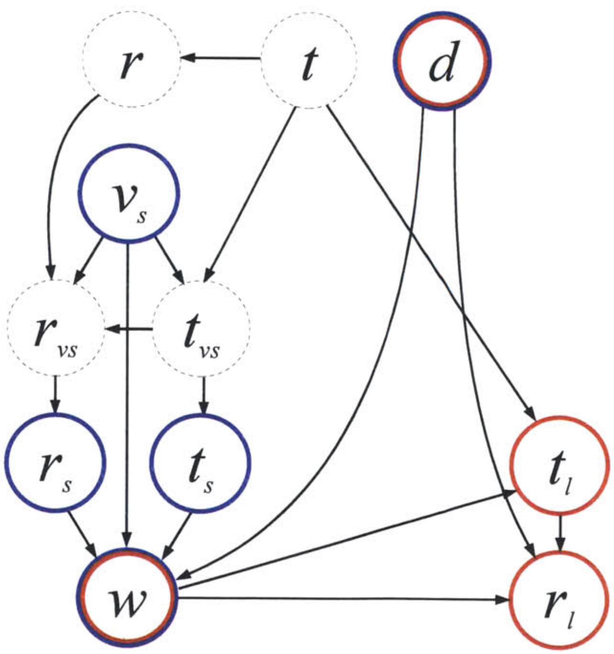
2.2.2 Speaker

The speaker has some noisy perception of the total set size and the reference set size, and has to choose a single word to say to the listener, to maximize the probability that the listener will correctly guess the *actual* reference set size (not the speaker's perceived reference set size). He considers a communication successful only if he communicates the exact quantity, so if the actual reference set size is 99, it is just as undesirable for the listener to guess 98 as to guess 2 (this assumption is for computational simplicity and could straightforwardly be generalized).

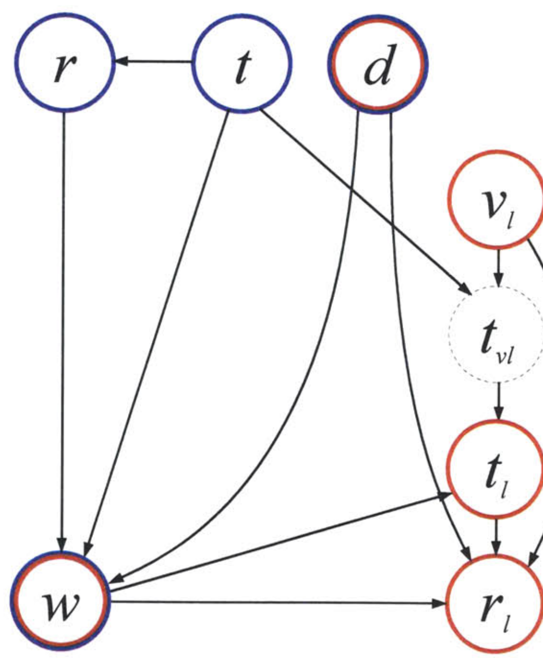
To choose a word, the speaker has to consider three things: his own perceptual uncertainty, the listener's perceptual uncertainty, and the method by which the listener will guess



(a) Full model.



(b) Speaker's model.



(c) Listener's model.

Figure 2-1: Directed graphical models. Blue-outlined variables are known to the speaker, and red-outlined variables are known to the listener.

a reference set size after hearing a word. To avoid infinite-recursive definitions (in which the speaker considers that the listener considers that the speaker considers that ...), the speaker assumes that the listener will guess a reference set size uniformly from all those of which the word is true in her perception of the world. He also assumes that although the listener has imperfect perception of the total set size, she will only consider total set sizes consistent with the presupposition of the word she has heard. However, he does *not* assume this about the word's assertion – for example, he assumes that if the listener only perceives a total set size of 1 but he says "two," then she cannot guess correctly, whereas if he says "both," she will revise her belief about the total set size and guess 2. This difference in treatment between the presupposition and assertion is motivated by computational simplicity but also seems not unreasonable for real communication.

For each word, the speaker computes the probability that the listener (or rather, his simplified approximation of the listener), after hearing that word, will guess the correct reference set size. This depends on two factors. The first factor is the reference set size: the speaker does not know the actual value r, so he must marginalize over all possible values. The second factor is, for each possible value of the actual reference set size r, the probability that the listener's guessed size r_ℓ will in fact equal r. Putting these two factors together, the speaker gets:

$$P(\text{correct}|w, r_s, t_s, v_s, d) = \sum_r P(r|r_s, t_s, v_s)P(r_\ell = r|w, r, t_s, v_s, d). \tag{2.1}$$

The probability of each actual reference set size can be computed using Bayes' Rule:

$$P(r|r_s, t_s, v_s) \propto P(r_s|r, t_s, v_s)P(r|t_s, v_s). \tag{2.2}$$

Now the speaker must consider his own perceptual uncertainty. If v_s is true (the set is completely visible), then the first term in (2.2), $P(r_s|r, t_s, v_s)$, is Gaussian with a known coefficient of variation, so it can be computed directly. If v_s is false, then the speaker must marginalize over the possible values of the actual total set size, the visible total set size, and the visible reference set size:

$$P(r_s|r, t_s, v_s) = \sum_t \sum_{t_{vs}} \sum_{r_{vs}} P(t|t_s, v_s)P(t_{vs}|t, v_s)P(r_{vs}|r, t_{vs}, v_s)P(r_s|r_{vs}, t_{vs}). \tag{2.3}$$

Similarly in the second term in (2.2), the speaker marginalizes over the possible values of the actual total set size, and applies Bayes' Rule:

$$P(r|t_s, v_s) = \sum_t P(r|t)P(t|t_s, v_s) \propto \sum_t P(r|t)P(t_s|t, v_s)P(t). \tag{2.4}$$

The speaker knows the distribution for the reference set size $P(r|t)$ (uniform) and the distribution for the total set size $P(t)$ (also uniform). The distribution of his observed total set size $P(t_s|t, v_s)$ can be computed analogously to $P(r_s|r, t_s, v_s)$ above: if v_s is true (the entire set is visible), then it is Gaussian; otherwise, the speaker must marginalize over the possible values of the visible total set size.

Returning now to the second term in equation (2.1), the probability of the listener's guessing a given reference set size r, the speaker must consider the listener's perceptual uncertainty. For simplicity he assumes that the set is completely visible to the listener, so he can ignore v_ℓ. The listener's guess depends both on the word w and on her perceived total set size t_ℓ, so the speaker must marginalize over values of t_ℓ (and therefore also t):

$$\begin{aligned} &P(r_\ell = r|w, r, t_s, v_s, d) \\ &\qquad = \sum_t \sum_{t_\ell} \mathbb{1}_{w\text{ applies}} P(t|t_s, v_s)P(t_\ell|t, w)P(r_\ell = r|w, r, t_\ell, d), \end{aligned} \tag{2.5}$$

where $\mathbb{1}_{w\text{ applies}}$ is equal to 1 if word w is true for the given r, t_ℓ, and d, and 0 otherwise. This equation represents the speaker's model of how the listener interprets a word. There is some total set size t; the speaker's belief distribution over this value, $P(t|t_s, v_s)$, has already been computed. The listener perceives the total set size to be t_ℓ, which depends on t in the same way that t_s does (in equation (2.4); the speaker assumes that v_ℓ is true). When the listener hears the word w, any values of t_ℓ that do not satisfy w's presupposition get their probabilities set to 0 (and the probabilities for the remaining values are renormalized). Finally, the listener assumes that the word is true, and (in the speaker's model) guesses a reference set size r_ℓ uniformly at random from all the values of which the word is true given her perception. So $P(r_\ell = r|r, w, t_\ell, d)$, the probability that she guesses the correct value r, is $1/|w|$, where $|w|$ is the number of values of which w is true given t_ℓ and d.

The *speaker's probability* for a word w in a given context will be defined as the expression from equation (2.1), $P(\text{correct}|w, r_s, t_s, v_s, d)$, normalized such that the speaker's probabilities for all words in a context sum to 1. In other words, the speaker's probability for a word is proportional to the probability that if he uses that word in that context, the (simplified) listener will guess the correct reference set size. Using these probabilities, the model has a parameter to choose between two options for how the speaker selects a word: either his probability of selecting a word can be proportional to its speaker's-probability, or he can choose the word with the maximum speaker's-probability.

2.2.3 Listener

The listener has some noisy perception of the total set size but not the reference set size, and hears a word from the speaker. She assumes that the speaker's word is true, and that the speaker chose the word to be informative – that is, that his probability of choosing a given word was inversely proportional to the number of possible reference set sizes of which it was true. She also assumes that the speaker has perfect perception of the world, i.e., that he knows the total set size t and the reference set size r exactly.

She first computes a distribution over the total set size, based on her perception and on the presupposition of the speaker's word (but not the assertion of the word). Then she computes the distribution over reference set sizes:

$$\begin{aligned} P(r|w, t_\ell, v_\ell, d) &= \sum_t P(t|t_\ell, v_\ell, w)P(r|w, t, d) \\ &\propto \sum_t P(t_\ell|t, v_\ell)P(t)P(w|r, t, d)P(r|t). \qquad (2.6) \end{aligned}$$

Most of the terms in (2.6) are analogous to those from the speaker's computation. The distribution $P(w|r, t, d)$ is the probability that the speaker (or rather, the listener's idealized version of him) would choose word w given r, t, and d, and is equal to $1/|w|$, where $|w|$ is the number of values of which w is true given t and d.

Chapter 3

Results

The model was evaluated using simulations and comparisons to existing languages. Parameters for perceptual uncertainty were chosen to approximate reality, though of course real-world conditions vary widely. Unless stated otherwise, the coefficient of variation for exact-number uncertainty was 0.2 (see, for example, Whalen et al., 1999), and the probability of the entire set being visible was 0.8. The total set size was drawn uniformly from 1 to 100.

3.1 Efficiency of Real Languages

If the model is assumed to be an approximately accurate representation of real communication with quantificational words, then it can be used to assess properties of existing languages. This analysis focuses not on the individual words but on the efficiency of communication using particular sets of words.

3.1.1 Empirical Data

In order to use the model on real languages, data had to be gathered about what quantificational words exist in what languages. A survey was conducted in which speakers of non-English languages reported the quantificational words that exist in those languages. They were given the following description:

> Quantifiers are, roughly, words that specify amounts – in English, they are approximately the words that can be answers to "how much"/"how many".
> Some examples of quantifier words in English are

'some', 'all', 'every', 'each', 'none', 'a', 'the', 'both', 'half', 'neither', 'either', 'any', 'few', 'several', 'enough', 'many', 'much', 'most', etc.
Some examples of quantifier meanings that English does not have single words for are things like
'all but one', 'less than half', 'some but not all', 'almost all', 'about how much you'd expect', etc.

They were then asked to report all of the quantifier words in their language, along with English translations.

There were a total of 45 respondents, of whom 5 were excluded for not following instructions. Some of the respondents were contacted with followup questions, to clarify their definitions and cross-check the responses of other people about the same language. The resulting data included 13 languages[1] (plus English).

Word	Number of Languages	Included in Model
a / one	10	✓
about half / a good amount	1	
all	14	✓
any	9	
both	12	✓
(a) couple	5	
every / each	14	(✓)
either	3	
(a) few	6	✓
half	14	✓
(a) little	10	(✓)
many	6	✓
most	11	✓
much	11	(✓)
neither	4	✓
no / none	10	✓
several	8	
some	14	✓
the	8	✓
very few	1	

Table 3.1: Frequency of quantificational words in 14 different languages. The words with checkmarks are the ones capable of being represented in the model. The words with parenthesized checkmarks cannot be exactly represented by the model, but can be considered the same as other words that can (for example, if a language has a word for "much" but not "many," for modeling purposes it is considered to have "many").

[1]Chinese, French, German, Haitian Creole, Hindi, Marathi, Polish, Portuguese, Romanian, Russian, Serbian, Spanish, and Yiddish.

Many respondents reported words that were questionably not (or clearly not) quantificational. The criteria for inclusion were that a word should be something that would be used to indicate how much or how many of something have some particular quality, if a situation were presented with no context. Thus words that depend on specific other context-dependent quantities, such as "enough" or "more," were not included.

Table 3.1 shows a list of the approximate meanings of the reported words that fit the criteria. Not all of these meanings are exactly the same as English's, however. For example, in some languages the word for "some" can also mean "few" or "several" or "any." The reason these meanings are not listed separately is that many respondents only gave single-word English translations even when the English word had a different meaning, so it would be inaccurate to count separately the words that did have more thorough translations.

It is also likely that some existing words were not reported. There is the concern that the instructions could bias respondents toward simply translating the English examples into the other languages; the inclusion of examples that do not have English words was intended to mitigate this factor. In addition, few respondents listed all of the words that were reported for their language, even words that *were* specifically given as examples in English. And for many of the languages, many of the meanings that were not reported were indeed confirmed not to have words.

The primary observation is that different languages do not seem to divide the space of meanings along different boundaries (as they do with, for example, color words). Rather, languages with fewer quantificational words tend to have a subset of the same meanings used in languages with more words. This is perhaps not surprising, given that unlike in domains such as color, there are few well-defined boundary points, and the meanings of the words overlap significantly. It should also be noted that even if the literal meanings are the same, the usages in practice could be substantially different.

3.1.2 Analysis of Sets of Meanings

For each of the languages, the set of its meanings representable by the model was analyzed. For comparison, other random sets of words were generated from a larger set consisting of all of the words from the real languages, plus other plausible words that were not reported in any of the languages ("some but not all," "less than half," "exactly the default-proportion," "not the" (i.e. "zero of one"), and "not all").

Across languages, the average proportion of words that are literally true in a situation is negatively correlated with the total number of quantificational words in the language ($r = -0.90$, $p < .01$) (Figure 3-1). In languages that have few quantificational words, a larger proportion of them tend to be true in each situation, whereas in languages with many quantificational words, a smaller proportion of them tend to be true. This result would be expected if languages with fewer words tend to have broader, more generally applicable words, and the additional words in bigger languages are specific to certain scenarios. In contrast, when sets of words are selected randomly, there is no correlation between the number of words and the average proportion that are true.

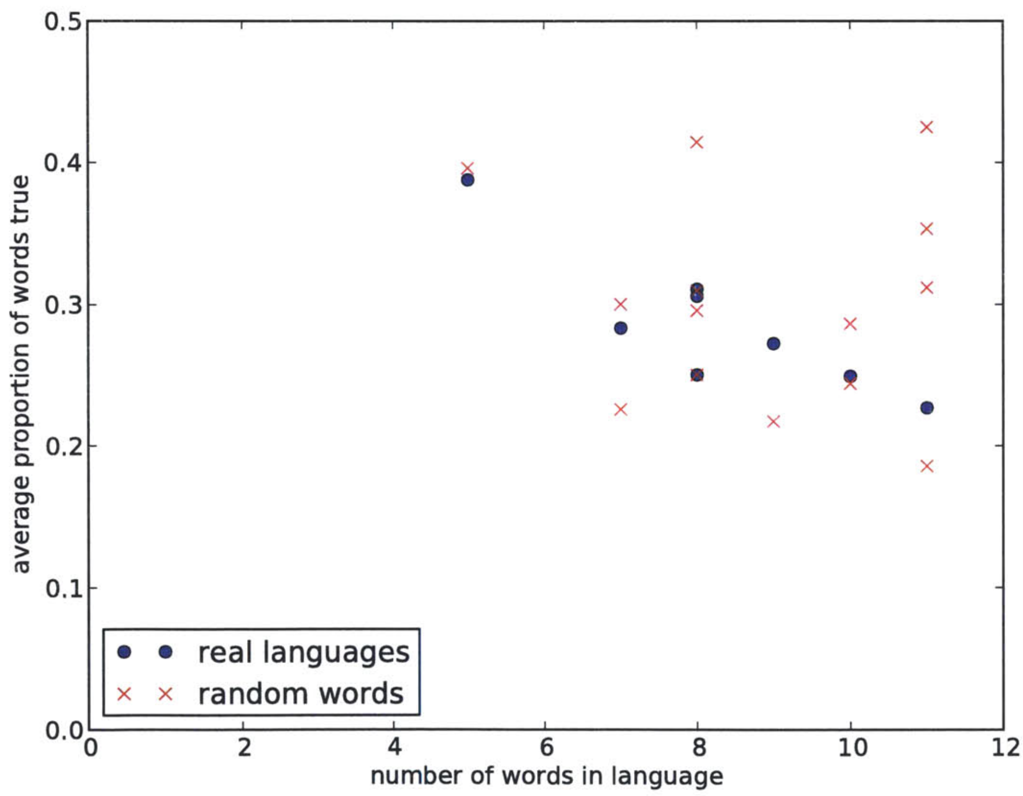


Figure 3-1: Average proportion of words true in a given situation, plotted against total words in the language, for real languages and for random sets of words.

To assess the communicative efficiency of sets of words in existing languages, simulations were run in which 1000 scenarios were randomly sampled for each language. For each scenario, the speaker chose a word with probability proportional to (his estimate of) the probability that the listener would guess the correct reference set size. Figure 3-2 plots the average proportional difference between the actual reference set size and the quantity that the listener guessed, for each of the languages. Languages with more words have lower average errors ($r = -0.92$, $p < .01$); this is not just a result of the informative communication assumption, because random sets of words did not show this correlation ($r = -0.34$, $p = .24$). The random sets of words also had an overall higher average error than the real languages ($t = -4.81$, $p < .01$).

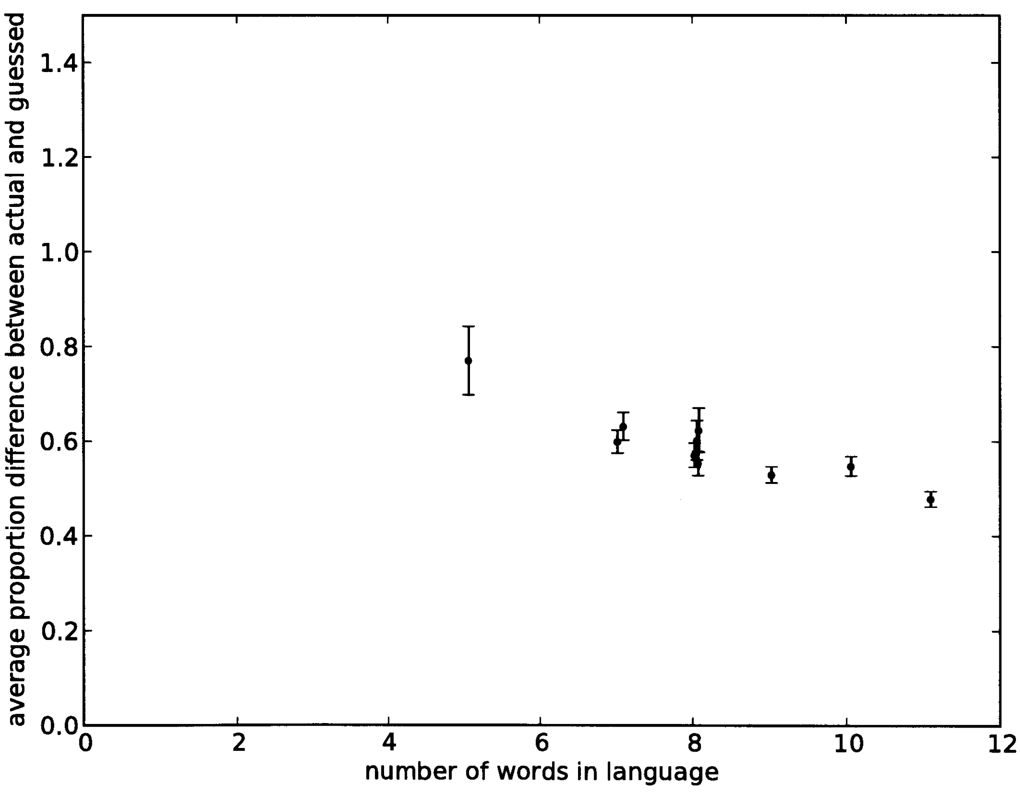


Figure 3-2: Average proportion difference between actual reference set size and listener's guessed reference set size, as a function of number of words in the language. Error bars show $\pm$ one standard error of the mean.

3.2 Usage Patterns

The model also makes specific predictions about how individual words will be used, depending on the context and on what other words are available. The accuracy of these predictions shows that the model is a reasonable representation of real-world communication of quantities.

3.2.1 Example Words

For a given set of words, the the speaker's probabilities across different situations qualitatively correspond to how the words are used in real communication. Figure 3-3 shows, for a particular set of seven words and a constant total set size, the speaker's probability for each word across different sizes of the reference set. Even though the literal meanings of the words are either true or false for a given reference set size, the probabilities are graded rather than all-or-nothing. It is also important to note that the probabilities are not simply how likely the speaker believes each word to be true: for example, "all" is true of only a single reference set size, yet the speaker gives it a high probability for all large reference set sizes because it would make the listener more likely to guess correctly.

This example for the most part accords with intuition about how these words are used in communication. The probabilities assigned by the model seem to approximate the appropriateness of each word. For small quantities, the speaker is relatively certain about the best words to use, and for larger quantities there are more words that could be used.

The results also bring to attention some unrealistic features of the model. One noticeable fact is that the speaker can assign a high probability to a word even if that word has a very low probability of actually being true. This can be viewed as a tradeoff between two factors: on the one hand there are precise words like "all," with a low probability of being true but a high probability of communicating an exact quantity; and on the other hand there are broad words like "some," with a high probability of being true but a low probability of communicating an exact quantity. In real-world communication, precision is somewhat important, but the model overemphasizes this factor because it treats correctness as a binary value. For the same reason, the word "some" in this example is not given as

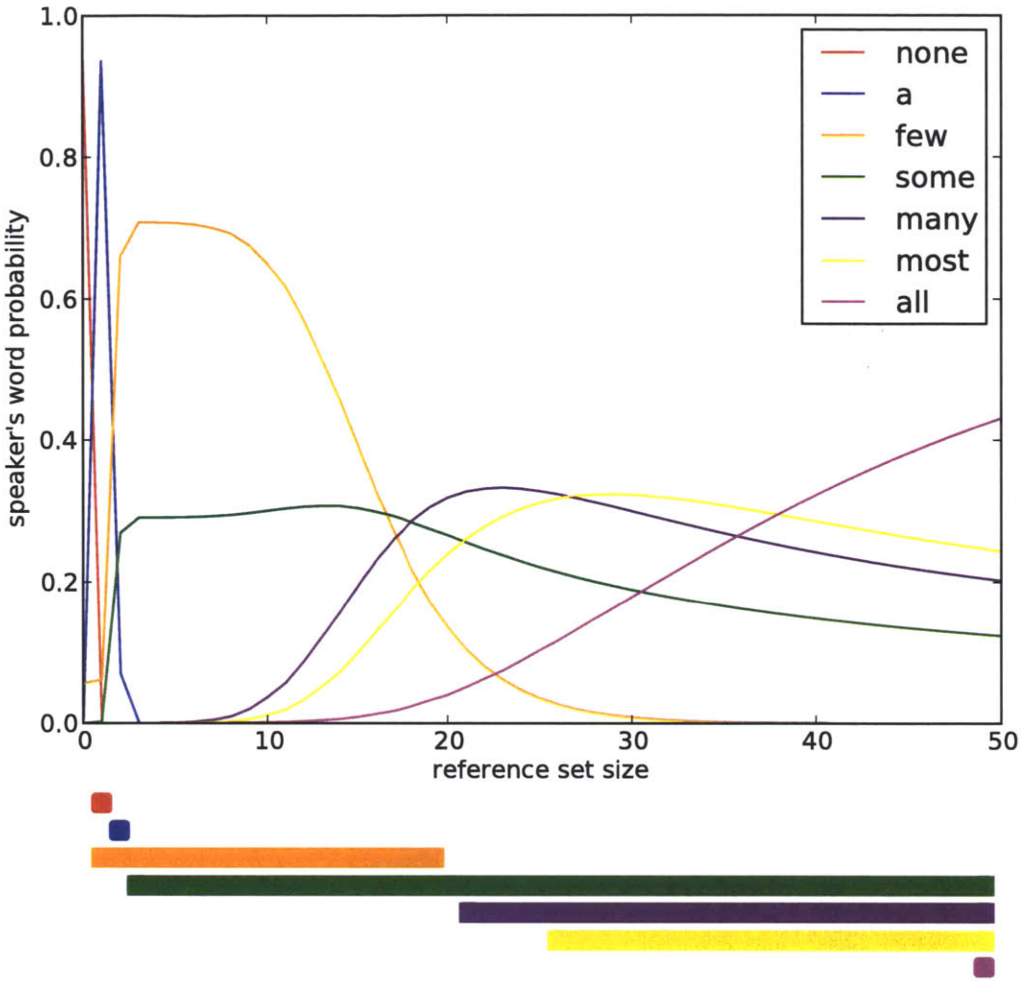
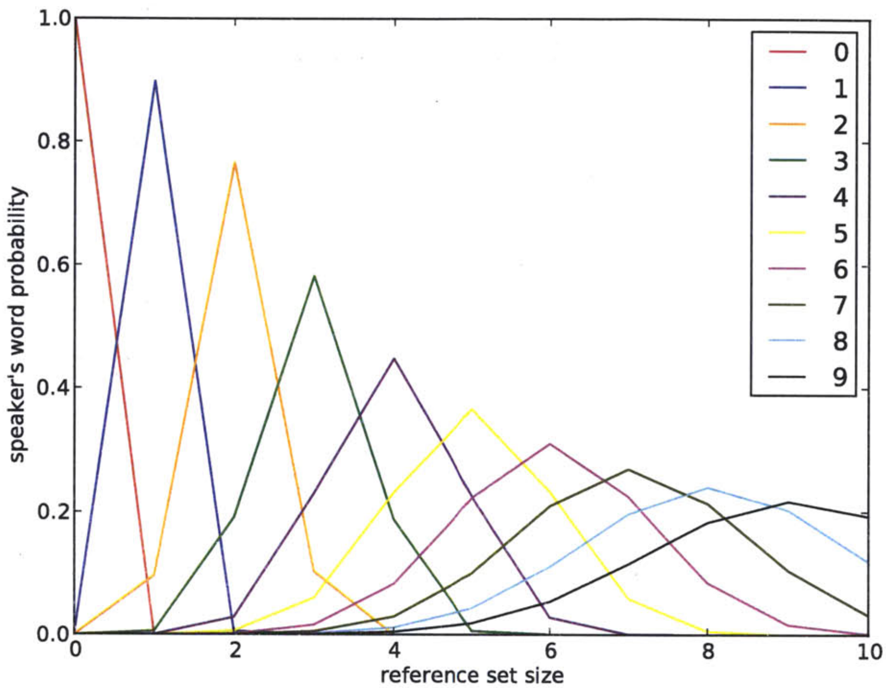


Figure 3-3: Speaker's probabilities for each word in an example vocabulary, over different reference set sizes, when the total set size is 50 and the default proportion is 0.4. The bars at the bottom show the reference set sizes of which each word is literally true.

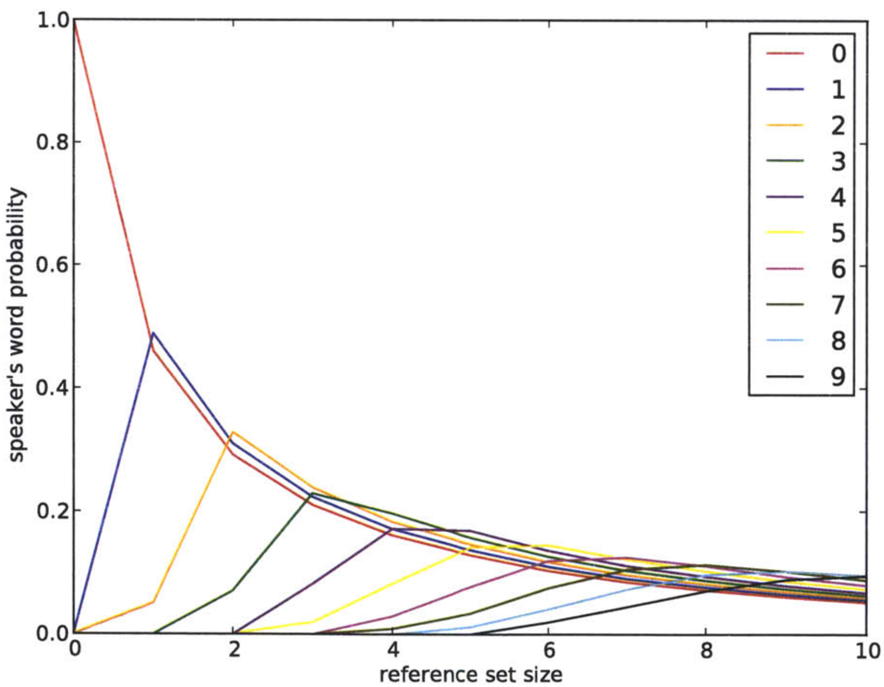
high of a probability as might be expected. Although "many" and "few" are more precise, real-world speakers are sometimes not concerned with the additional information that those words carry, and might simply want to convey the concept of "some."

3.2.2 Numbers

Number words were not included in the sets of words for real languages because all of the languages from the survey have effectively infinitely many words for integers. However, the model can be used just as well for numbers, and they provide a particularly clean example that demonstrates the interaction between informativeness and perceptual uncertainty.



(a) Exact number meanings.



(b) Lower-bounded number meanings.

Figure 3-4: Speaker's probabilities for (a) exact meanings of number words and (b) lower-bounded meanings of number words (e.g. "at least three"), over different reference set sizes (the total set size is 20).

Figure 3-4 shows the speaker's probabilities for a vocabulary with lower-bounded number meanings (such as "at least three") versus a vocabulary with exact number meanings. For smaller quantities, the distributions are very peaked around the exact reference set size that the speaker perceives: he is fairly sure of the quantity, and consequently fairly sure that the listener will guess correctly if he uses that number word (and not if he uses another word). As the numbers get larger, the distributions get flatter because of perceptual uncertainty. If the coefficient of variation for perceptual uncertainty is set to a lower value (corresponding to, for example, a person explicitly counting), then the speaker's probabilities become much more peaked around the exact quantities.

For the lower-bounded meanings the same general pattern is visible. Words with lower bounds closer to zero are more likely to be true, but they are less informative because there are more possible reference set sizes of which they are true. The results for the lower-bounded meanings demonstrate a case where an additional level of recursion might be useful. In the current model, the speaker assumes that the listener will guess uniformly from all reference set sizes of which the word is true, and so there is not much difference in informativeness between "at least n" and "at least $n + 1$" (especially if the total set size is large). If the speaker knew that the listener was assuming that he was speaking informatively, this difference would be more important.

3.2.3 Presupposition

In this model, the purpose of a word's presupposition is for the speaker to provide information about the total set size, which can be useful because of the listener's perceptual uncertainty. The model demonstrates that even without an explicit "maximize presupposition" maxim, presuppositions will naturally be maximized by a speaker trying to communicate informatively to a listener with imperfect perception. Figure 3-5 shows the speaker's probability for numerical words with and without presuppositions (such as "two" versus "both"), compared with the word "all." As the quantities get bigger, the presupposition becomes less useful, because it is less likely to be true. This result is consistent with the observation that in the survey data, all of the words that have presuppositions presuppose a total set size of 1 or 2, never any larger numbers.

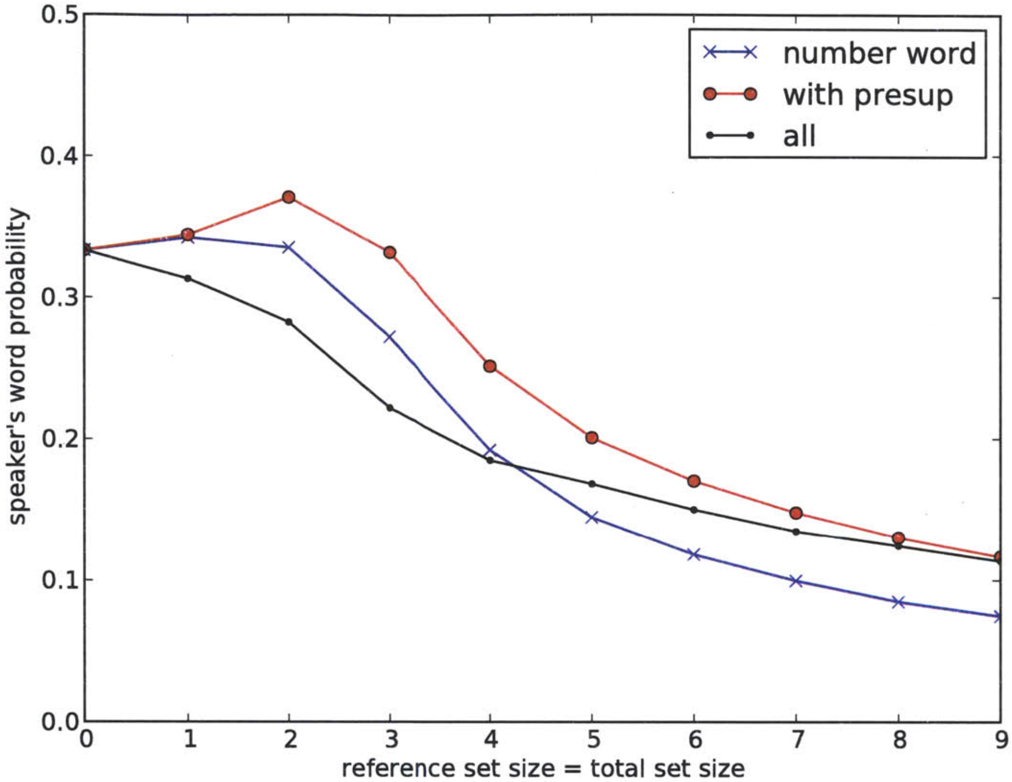


Figure 3-5: Speaker's probability for exact number words ("n of the things") versus number words with presupposition ("n of the n things") versus the word "all," for scenarios in which the reference set size is equal to the total set size.

Similar to the example in Figure 3-3, the model does seem to care too much about exact quantity, and the relative probabilities of "all" compared with the presuppositional words are lower than one might expect. Another fact not captured here is that although words with stronger presuppositions might be more informative, they are applicable in fewer situations. It might therefore not be worthwhile for a language to give them their own words, because in the few situations where they would be used, a combination of other words could be used instead (for example, "all n" or "the n" in English).

Chapter 4

Conclusion

4.1 Summary

This thesis has presented a new framework for modeling linguistic communication of quantities. The model uses a simple representation of word meanings as intervals that can depend on both relative and absolute quantities, and includes realistic assumptions about communicating informatively in the presence of perceptual uncertainty. Each word is either true or false in a given situation, but the model predicts that some words are more likely than others to be used in certain situations, just as is observed in real-world communication. Furthermore, when the sets of words from real languages are used by the model, they tend to result in lower error rates for communication than sets of words selected randomly.

4.2 Possible Extensions

4.2.1 Additional Features

Some of the features in the meanings of real quantificational words and communicative situations are not currently represented in this model but could be incorporated into it.

The communicative intention of the speaker is quite simplified in the basic model. In the real world, a speaker often does not care about the exact quantity, and in fact sometimes there is not any exact quantity. A speaker might want to convey only a particular aspect of the quantity (for example, that it is larger than expected) and be satisfied if the listener inferred any quantity in some range. Or a speaker might want to convey a particular distribution over quantities. It would be straightforward to modify the model such that the speaker considered a communication equally successful if the listener guessed any quantity within some range, or such that the speaker tried to communicate a specific belief distribution over quantities instead of having the listener guess any single quantity.

The model also does not include any representation of subjective "degree," such as the difference between "many" and "very many." This is because in any particular communicative context it only has one "default" proportion. It would be possible to instead keep track of multiple default proportions, a different one for each word that needed one, so that "many" could be represented as "more than Default Proportion x" and "very many" could be represented as "more than Default Proportion y". It could also be useful to have default absolute amounts in addition to default proportions.

Relatedly, the meanings of some real words depend on both relative quantities *and* absolute quantities. For example, it would sound odd to use "many" to refer to two out of three, because "many" generally has to be a large absolute amount in addition to being a large proportion. Such meanings could be represented in the model by allowing the interval representation of a meaning to have multiple lower and upper bounds, so that the meaning only applied to quantities that were greater than all of its lower bounds and less than all of its upper bounds.

4.2.2 Additional Experiments

The model's results are consistent with general intuition about the the meanings and usage of quantificational words; however, it will be useful to quantitatively compare the predictions with real speakers' probabilities of using words and real listeners' probabilities of guessing quantities. These probabilities could easily be obtained by presenting people with the same controlled contexts used in the model, and either asking for probabilities directly or asking for a single choice selected from a list of options and computing the probabilities across participants.

It would also be interesting to conduct a more thorough investigation of what quantificational words exist in what languages. To address the possible English-bias of the survey from Chapter 3, an experiment could be conducted in which words are elicited not by giving English examples but by displaying actual scenarios. This method would provide more precise information about the meanings, and could be structured to determine each word's literal meaning as well as in what contexts it is actually used.

Overall, the model presented here has provided some interesting insights into the workings of quantificational words, and has set up a foundation for further investigations.

References

Bach, E., Jelinek, E., Kratzer, A., & Partee, B. H. (Eds.). (1995). *Quantification in natural languages.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, *4*(2), 159-219.

Bass, B. M., Cascio, W. F., & O'Connor, E. J. (1974). Magnitude estimations of expressions of frequency and amount. *Journal of Applied Psychology*, *59*(3), 313-320.

Chase, C. I. (1969). Often is where you find it. *American Psychologist*, *24*(11), 1043.

Coventry, K. R., Cangelosi, A., Newstead, S., Bacon, A., & Rajapakse, R. (2005). Grounding natural language quantifiers in visual attention. In *Proceedings of the 27th annual meeting of the cognitive science society.*

Coventry, K. R., Cangelosi, A., Newstead, S. E., & Bugmann, D. (2010). Talking about quantities in space: vague quantifiers, context and similarity. *Language and Cognition*, *2*, 221-241.

Dehaene, S. (1999). *The number sense: How the mind creates mathematics.* USA: Oxford University Press.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology*, *46*, 621-646.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998.

Frank, M. C., Goodman, N. D., Lai, P., & Tenenbaum, J. B. (2009). Informative communication in word production and word learning. In *Proceedings of the 31st annual meeting of the cognitive science society.*

Geurts, B., Katsos, N., Cummins, C., Moons, J., & Noordman, L. (2009). Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes*, *25*, 130-148.

Gierasimczuk, N. (2007). The problem of learning the semantics of quantifiers. *Logic, Language and Computation*, *4363*, 117-126.

Gil, D. (2001). Quantifiers. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals* (p. 1275-1294). Berlin; New York: Walter de Gruyter.

Goodman, N. D., & Stuhlmüller, A. (2012). Knowledge and implicature: Modeling language understanding as social cognition. In *Proceedings of the 34th annual meeting of the cognitive science society.*

Graff, P., Romoli, J., Moro, A., & Snedeker, J. (2009). *Exploring a learning bias against non-conservative determiners.* Presented at the 22nd Annual Meeting of the CUNY Conference on Human Sentence Processing.

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

Hanlon, C. (1978). The emergence of set-relational quantifiers in early childhood. In

P. S. Dale & D. Ingram (Eds.), *Child language – An international perspective* (p. 115-130). Baltimore: University Park Press.

Hanlon, C. (1981). Frequency of usage, semantic complexity, and the acquisition of set-relational quantifiers in early childhood. In C. E. Johnson & C. L. Thew (Eds.), *Proceedings of the second international congress for the study of child language* (p. 245-260). Washington, D.C.: University Press of America.

Heim, I. (1991). Artikel und definitheit. In A. V. Stechow & D. Wunderlich (Eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung.* Berlin: de Gruyter.

Hunter, T., & Lidz, J. (submitted). *Some (but not all) unattested determiners are unlearnable.*

Keenan, E. L., & Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy*, *9*, 253-326.

Khetarpal, N., Majid, A., & Regier, T. (2009). Spatial terms reflect near-optimal spatial categories. *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*

Matthewson, L. (1996). *Determiner systems and quantificational strategies: Evidence from Salish.* Unpublished doctoral dissertation, University of British Columbia.

Moxey, L. M., & Sanford, A. J. (1993a). *Communicating quantities: A psychological perspective.* Hove, UK; Hillsdale USA: Lawrence Erlbaum Associates.

Moxey, L. M., & Sanford, A. J. (1993b). Prior expectation and the interpretation of natural language quantifiers. *The European journal of cognitive psychology*, *5*, 73-91.

Moxey, L. M., & Sanford, A. J. (2000). Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology*, *14*, 236-255.

Newstead, S. E., Pollard, P., & Riezebos, D. (1987). The effect of set size on the interpretation of quantifiers used in rating scales. *Applied Ergonomics*, *18*(3), 178-182.

Papafragou, A., & Musolino, J. (2002). The pragmatics of number. *Proceedings of the 24th Annual Conference of the Cognitive Science Society.*

Pearl, J. (1998). Graphical models for probabilistic and causal reasoning. In D. M. Gabbay (Ed.), *Handbook of defeasible reasoning and uncertainty management systems: Quantified representation of uncertainty and imprecision* (Vol. 1, p. 367-389). Springer.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2010). Beyond Boolean logic: exploring representation languages for learning complex concepts. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*(4), 1436-1441.

Schlenker, P. (to appear). *Maximize Presupposition* and Gricean reasoning. *Natural Language Semantics.*

Sullivan, J., & Barner, D. (2011). Number words, quantifiers, and principles of word learning. *Wiley Interdisciplinary Reviews: Cognitive Science.*

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, *115*(4), 348-365.

Weber, E. H. (1846). Tastsinn und Gemeingefühl. In R. Wagner (Ed.), *Handwörterbuch der Physiologie* (Vol. iii, p. 481-588). Brunswick: F. Vieweg und sohn.

Whalen, J., Gallistel, C., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*(2), 130-137.