Pragmatic factors can explain variation in interpretation preferences for quantifier-negation utterances: A computational approach

Noa Attali
UC Irvine
nattali@uci.edu

Gregory Scontras
UC Irvine
g.scontras@uci.edu

Lisa S. Pearl
UC Irvine
lpearl@uci.edu

Abstract

Traditional investigations of quantifier-negation scope ambiguity (e.g., *Everyone didn't go*, meaning that *no one went* or *not everyone went*) have focused on universal quantifiers, and how ambiguity in interpretation preferences is due to the logical operators themselves and the syntactic relation between those operators. We investigate a broader range of quantifiers in combination with negation, observing differences in interpretation preferences both across quantifiers and also within the same quantifier (confirmed by corpus analysis). To explain this variation, we extend a computational cognitive model that incorporates pragmatic context-related factors, and which previously accounted for *every*-negation, to predict human interpretation preferences also for *some* and *no*. We evaluate the model's predictions against human judgments for quantifier-negation utterances, finding a strong qualitative and quantitative match when the listener has particular expectations about the world in which the utterance occurs. These results suggest that pragmatic factors can explain variation in interpretation preferences.

Keywords: ambiguity resolution; scope; computational models

Introduction

Utterances like (1) that involve a quantified subject and sentential negation (i.e., **quantifier-negation** utterances) are reported to be ambiguous, allowing both a surface interpretation, (1a), and an inverse one, (1b), depending on the logical scope of the quantifier relative to negation (e.g., Musolino, 1999).

(1) Every horse didn't jump over the fence.
   a. No horses jumped. *(every>n't)*
   Surface scope: $\forall$x[horse(x) $\rightarrow$ $\neg$jump(x)]
   b. Not all the horses jumped. *(n't>every)*
   Inverse scope: $\neg\forall$x[horse(x) $\rightarrow$ jump(x)]

Given the structural mechanism that delivers the two interpretations (e.g., "quantifier raising"; May, 1977), we might expect that identically-structured utterances should be similarly ambiguous. That is, if the particular logical operators (here: universal *every* and negation *n't*) and the syntactic relation between them are the most important factors determining scope interpretation, we might expect similar interpretation preferences for the same quantifier-negation configurations. However, the existing experimental literature claims that native English speakers exhibit a range of interpretation preferences for utterances like (1) with universally-quantified subjects (i.e., *every*, *all*): sometimes they prefer inverse scope (Carden, 1973; Musolino et al., 2000; Musolino and Lidz, 2006), sometimes surface (Conroy et al., 2008; Lee, 2009), and sometimes they show no preference (Lee, 2009).

Examples (2)-(3) further illustrate potential variation in interpretation preferences for the same quantifier-negation configuration. Both (2) and (3) feature universal *every*, yet our intuitions suggest that the inverse interpretation is preferable in (2), while the surface interpretation seems preferable in (3) (preferred interpretations **bolded**).

(2) Everyone isn't going to college.
   a. No one is going to college. *(every > n't)*
   b. **Not everyone is going to college.** ***(n't > every)***

(3) Everyone isn't going to live forever.
   a. **No one is going to live forever.** ***(every > n't)***
   b. Not everyone will live forever. *(n't > every)*

Our world knowledge about probable interpretations appears to help disambiguate (2) and (3). In (2), we know that many people do attend college, so the surface interpretation is unlikely to be true. In (3), our knowledge of human biology tells us that the surface interpretation is highly likely to be true.

Beyond variation within the same quantifier (i.e., within *every*-negation constructions), there is also variation across quantifiers: that is, different quantifiers privilege different interpretations in quantifier-negation utterances. For instance, in (1), *every* seems relatively ambiguous in combination with negation, with inverse scope perhaps preferred. In contrast, existential *some* in (4) and *no* in (5) seem to favor the surface interpretation: (4) seems to do so strongly while the preference in (5) may be less obvious.

(4) Some horse didn't jump over the fence.
   a. **There's a horse that didn't jump.** ***(some > n't)***
   Surface scope: $\exists$x[horse(x) & $\neg$jump(x)]
   b. There are no horses that jumped. *(n't > some)*
   Inverse scope: $\neg\exists$x[horse(x) & jump(x)]

(5) No horse didn't jump over the fence.
   a. **There isn't a horse that didn't jump.** ***(no > n't)***
   Surface scope: $\nexists$x[horse(x) & $\neg$jump(x)]
   b. There aren't zero horses that jumped. *(n't > no)*
   Inverse scope: $\neg\nexists$x[horse(x) & jump(x)]

Taken together, these observations about variation in the interpretation preferences of quantifier-negation utterances suggest that preferences depend on (i) the structural configuration of logical operators (i.e., the quantifier-negation construction itself, which licenses ambiguity), (ii) the lexical content (i.e.,

the specific quantifiers involved), and (iii) the broader context (e.g., our world knowledge about which interpretations are more or less likely). We aim to show how probabilistic models of scope ambiguity resolution that take all these factors into account—particularly context—stand the best chance of explaining within-quantifier and across-quantifier variation.

We first seek to confirm within-quantifier variation of interpretation preferences in context by examining naturally-occurring *every*-negation utterances in the Corpus of Contemporary American English (Davies, 2015) and crowd-sourcing human interpretation preferences for these utterances. We indeed find a good deal of variation: some *every*-negation utterances show a strong preference for surface interpretations, others show a strong preference for inverse interpretations, and others remain ambiguous. We then review how a computational cognitive model of scope ambiguity resolution proposed by Savinelli et al. (2017), which incorporates the pragmatic factors of world expectations and conversational goals, could explain the variation we document. Next, we extend this model beyond *every*-negation utterances to explore the lexical source of variation introduced by different quantifiers, namely *some* and *no*. We assess the predictions of the extended model against human interpretation preferences for *some*-negation and *no*-negation utterances; we find that the model's predictions have both a good qualitative and quantitative fit, but only when the listener has particular pragmatic expectations. We discuss the implications of our findings for theories of quantifier-negation ambiguity representation, and for the pragmatics of ambiguity resolution more broadly.

Human-annotated corpus analysis

Corpus search of *every*-negation

We extracted 390 instances of *every*-negation utterances in the speech genre of the Corpus of Contemporary American English (**COCA**), where quantified subjects precede and c-command sentential negation (*not* or contracted *n't*). COCA contains transcripts of spoken conversations from American radio and TV programs from 1990 to 2012 (≈9 million clauses).

Crowd-sourcing interpretation preferences

Following Degen (2015), we annotated ambiguous utterances with their interpretations by asking participants to rate these utterances in their context. Interpretations were measured on a sliding scale following the paraphrase-endorsement methodology of Scontras and Goodman (2017), described below.

Participants. We recruited 150 participants with U.S. IP addresses through Amazon.com's Mechanical Turk (MTurk) crowd-sourcing service. Of these 150, we assess data from 48 participants (43% female; mean age: 41) who passed attention and understanding controls and indicated that English was their only native language. Each received $2.00.

Design. Participants were asked to "choose the best paraphrase for the bolded part" for fifteen randomly-selected conversation excerpts (see Figure 1). Excerpts consisted of three preceding sentences, a single bolded potentially-ambiguous clause, and one following sentence. Beneath the excerpt, participants rated paraphrases of the surface and inverse scope interpretations on a scale between "definitely not" and "definitely". Because the ambiguous clauses took the form *quantified noun phrase–negation–verb–remainder* (e.g., *Everybody's not doing it*), surface scope paraphrases took the form *none/no one/nobody/nothing–verb–remainder* (e.g., *nobody is doing it*) and inverse scope paraphrases took the form *not all/not all things–remainder* (e.g., *not all are doing it*).

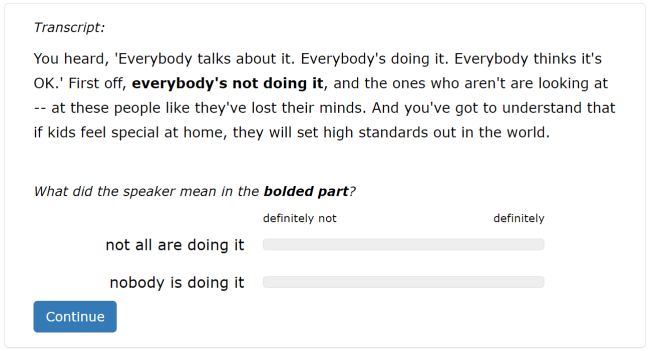

Figure 1: Sample paraphrase-endorsement trial from the crowd-sourced corpus analysis.

Results. We report preliminary results from the ratings by the 48 participants who passed all controls; limiting ourselves to items with at least two different participant ratings, this process yielded scores for 223 of the 390 utterances. Judgments tended to show a negative correlation between the agreement with the surface and inverse scope paraphrase, with few judgments showing strong agreement or disagreement with both paraphrases (see Figure 2).

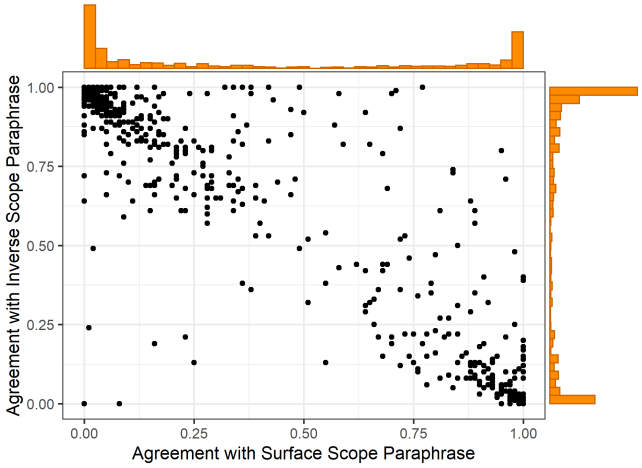

Figure 2: Individual participant endorsement ratings of the surface and inverse paraphrases of individual items (r=-.93).

Given this general negative correlation, we decided to use the mean difference between participant endorsement of the surface and inverse scope paraphrases (see Figure 3) as the measure of an item's preferred interpretation. This mean difference ranges between -1 (maximum endorsement of inverse in-

terpretations, with surface paraphrases rated as "definitely not" [0] and inverse paraphrases as "definitely" [1]: $0-1=-1$) and 1 (maximum endorsement of surface interpretations, with surface paraphrases rated as "definitely" [1] and inverse paraphrases as "definitely not" [0]: $1-0=1$).

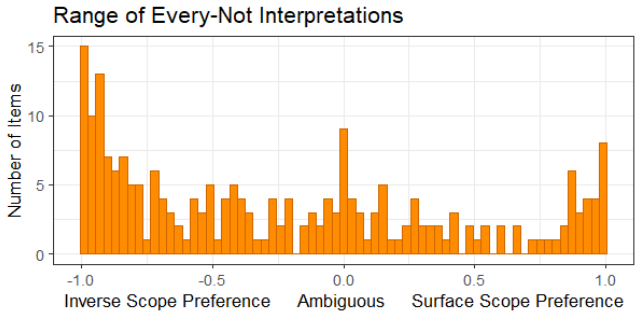

Figure 3: Mean difference between surface and inverse scope interpretation agreement per item in the corpus analysis.

Figure 3 shows the wide range of interpretation differences for the *every*-negation utterances in our preliminary corpus analysis. This evidence suggests that many naturally-occurring potentially-ambiguous utterances elicit strong, reliable intuitions such that they are indeed unambiguous in context: 43% of our sample had scores below -0.5 (strongly inverse) while 17% had scores above 0.5 (strongly surface). More importantly, we see that different *every*-negation utterances receive different interpretations in context. These within-quantifier interpretation preferences must rely on factors beyond the quantifier semantics or quantifier-negation configuration, since these elements are the same in our sample.

We illustrate some of the observed variation with examples of a strong surface scope preference in (6) (mean difference≈1.0), a strong inverse scope preference in (7) (mean difference≈-1.0)), and true ambiguity in (8) (mean difference≈0.0). For each example, we report the number of participants who judged the item, the mean difference score, and the standard deviation.

(6) *Surface*: I wanted to hear like a real type of radio. Everybody was milquetoast. Everybody was middle of the road. **Everybody didn't want to offend everybody else** [N = 5, M = 0.97, SD = 0.06] and it was boring.

(7) *Inverse*: It won't happen in a day, but people should start talking and taking hands and say now we know this happened but **everyone on campus isn't a racist** [N = 3, M = -0.93, SD = 0.05]. **Everyone doesn't hate black people** [N = 3, M = -0.99, SD = 0.02], **every black person doesn't hate white people** [N = 4, M = -1.00, SD = 0.005].

(8) *Ambiguous*: Shoshanna, how were you treated? JOHNSON: Surprisingly humanely. It wasn't perfect. **Everything wasn't kind and, you know, sweet or anything like that** [N = 6, M = -0.07, SD = 0.81], but during captivity, the worst things come to mind.

Model

Our hypothesis about the influence of pragmatic factors on interpretation preferences is formalized as an extension of the proposal by Savinelli et al. (2017). They use a Rational Speech Act (**RSA**) model (Frank and Goodman, 2012) to formally articulate the cognitive process that yields observed interpretations of scopally-ambiguous utterances. In particular, a listener's expectations about the world and the speaker's conversational goal are salient pragmatic components that make different interpretations more (or less) informative. So, considering the informativity of an interpretation in context could cause that interpretation to be viewed as more (or less) likely. Importantly, these pragmatic factors can vary both within and across quantifiers and so potentially lead to the variation we observe. We adapt Savinelli et al.'s model, which describes scope ambiguity resolution for *every* and negation, to additionally account for quantifiers *some* and *no*.

Here, we follow Savinelli et al. (2017) and assume that a speaker chooses from a set of utterances U which consists of the particular quantifier-negation utterance or saying nothing at all (*null*): {*every*/*some*/*no*-negation, *null*}.[1] In this way, the different quantifiers rely on different utterance alternatives, and so we avoid claiming that the three quantifiers are necessarily alternatives to each other (as they would be if U ={*every*-negation, *some*-negation, *no*-negation}).

For the utterance context, speakers describe a scenario with three marbles such that the possible world states are the number of marbles that are red: $w \in W = \{0,1,2,3\}$. Speakers and listeners know the utterance semantics to be a mapping, parameterized by the scope interpretation $i \in I = \{surface, inverse\}$, from world states $w \in W$ to truth values $Bool = \{\texttt{true}, \texttt{false}\}$. In other words, speakers and listeners share a common semantics for the utterances (9), which determines which states are `true` for a given interpretation. For example, the *every*-negation utterance (i.e., *Every marble isn't red*) maps world {0} to `true` under surface scope and worlds {0,1,2} (w≠3) to `true` under inverse scope.

(9) Utterance semantics $[\![u]\!]^i$:

a. $[\![\textit{every}\text{-negation}]\!]^{surface} = \lambda w.\ w = 0$
b. $[\![\textit{every}\text{-negation}]\!]^{inverse} = \lambda w.\ w \neq 3$
c. $[\![\textit{some}\text{-negation}]\!]^{surface} = \lambda w.\ w \neq 3$
d. $[\![\textit{some}\text{-negation}]\!]^{inverse} = \lambda w.\ w = 0$
e. $[\![\textit{no}\text{-negation}]\!]^{surface} = \lambda w.\ w = 3$
f. $[\![\textit{no}\text{-negation}]\!]^{inverse} = \lambda w.\ w > 0$
g. $[\![\textit{null}]\!] = \lambda w.\ \texttt{true}$

The speaker's conversational goal is to address the topic of conversation by guiding the listener to a set of intended world states. This set is determined by the question under discussion

[1] Savinelli et al.'s model is meant to capture truth-value judgments, where participants choose between endorsing or not endorsing the ambiguous utterance—hence the forced choice between the utterance and saying nothing (i.e., *null*). See Scontras and Goodman (2017) for an alternative model of ambiguity resolution that models paraphrase-endorsement data, as we do here. Future work can incorporate additional alternative utterance sets and explore their impact on predicted interpretation preferences.

(**QUD**). For instance, the QUD *all?* indicates that a speaker wants to resolve whether all the marbles are red (w$\in\{3\}$) or not (w$\in\{0,1,2\}$). The model implements a QUD as a mapping from worlds to partitioned sets of worlds x, as in (10); the full set of QUDs $q \in Q = \{$*all?, none?, how-many?*$\}$.

(10) QUD semantics $[\![q]\!]$:

a. $[\![$*all?*$]\!] = \lambda$w. w = 3

b. $[\![$*none?*$]\!] = \lambda$w. w = 0

c. $[\![$*how-many?*$]\!] = \lambda$w. w

The RSA model implements a series of recursive reasoning layers, with a speaker choosing utterances by reasoning about how a listener would interpret them, and a listener interpreting utterances by reasoning about the speaker who generated them. Here, a pragmatic listener L_1 reasons about the speaker S_1 who generated the utterance, and imagines that S_1 was reasoning about a literal listener L_0 when generating that utterance. The hypothetical literal listener L_0 hears an utterance u and knows its intended interpretation i; L_0 then reasons that the true state of the world w is any of the world states that are true, given the semantics of the ambiguous utterance $[\![u]\!]^i$ from (9). The model implements this reasoning as a filter on the possible world states $\delta_{[\![u]\!]^i(w)}$, which returns 1 when $[\![u]\!]^i(w)$ is `true` and 0 otherwise. L_0 then weights the possible worlds by L_0's prior beliefs about their probabilities $P(w)$.

$$P_{L_0}(w|u,i) \propto \delta_{[\![u]\!]^i(w)} \cdot P(w) \tag{11}$$

L_0 takes the QUD q into account by inferring the intended set of world states x determined by $[\![q]\!](w)$, as in (12). The model implements this inference via the filter $\delta_{x=[\![q]\!](w)}$, which is 1 when $x = [\![q]\!](w)$ and 0 otherwise.

$$P_{L_0}(x|u,i,q) \propto \sum_w \delta_{x=[\![q]\!](w)} \cdot P_{L_0}(w|u,i) \tag{12}$$

The speaker S_1 selects u, knowing the particular intended world w, scope interpretation i, and QUD q as in (13). This calculation is based on the the perceived utility of u, which depends on the probability of u communicating the intended set of world states x to L_0. This decision process is mediated by a softmax function and free parameter α which controls how the speaker perceives the relative contrasts between potential options; contrasts can be sharpened ($\alpha > 1$), smoothed away ($\alpha < 1$), or perceived as is (α=1).

$$P_{S_1}(u|w,i,q) \propto exp(\alpha \cdot \log(P_{L_0}(x|u,i,q))) \tag{13}$$

Hearing a quantifier-negation utterance, a pragmatic listener L_1 reasons jointly about the true world state w, scope interpretation i, and QUD q that would have been most likely to lead S_1 to produce the utterance that was observed. L_1 considers the prior probabilities of w, i, and q as well, as shown in (14).

$$P_{L_1}(w,i,q|u) \propto P(w) \cdot P(i) \cdot P(q) \cdot P_{S_1}(u|w,i,q) \tag{14}$$

Parameter setting. Savinelli et al. focused on modelling how child interpretation behavior differs from adult behavior for *every*-negation utterances in context; they found that one key factor for adult-like interpretations is a prior over world states that favors the *all* state (e.g., $w = 3$ in our 3-marble world) . We term this belief a "high positive expectation".

Our corpus data support the plausibility of a high positive expectation: many *every*-negation items are echoic, explicitly repeating and denying previous content that asserts a high positive expectation. Some examples of this use appear in (15)-(18), with the content causing the high positive expectation in ***bold italics*** and the *every*-negation utterance in **bold**.

(15) Mr. DEITCHMAN: ***Everybody on 86th Street here is young***, so they all enjoy it, they all like it.
STOSSEL voice-over: Well, **everyone on the street isn't young** and lots of people don't like it.

(16) Dr. SHALALA: We do have a health care crisis in this country and ***every American knows it***.
LIMBAUGH: No, **every American doesn't know it**

(17) I don't think you should allow ***every doctor to do it***—that's just my personal opinion—because **every doctor can't do it**, by religion or philosophy or personality.

(18) MICHAEL: I went back there and knocked on the door to see if ***everything was OK***.
RIVERA: Mm-hmm.
MICHAEL: And Mama said **everything wasn't OK**, so I called the police...

Motivated by Savinelli et al.'s analytic results and our impressions from the corpus, we set a high positive expectation in our model using Savinelli et al.'s parameter values ($P(w = 3) = 0.9$). We also followed Savinelli et al. by keeping uniform priors over QUD and scope interpretation. We set $\alpha = 1$ (i.e., the speaker perceives probabilities as is, neither sharpening nor smoothing away relative contrasts.)

Predictions. Under these parameter settings, the model predicts that the proportion of inverse interpretations depends on the quantifier, with *every*-negation most likely to receive an inverse interpretation, then *no*-negation, and then *some*-negation (see model predictions in Figure 8). Savinelli et al. showed how utterance informativity is a driving factor in model behavior, especially in cases of within-quantifier variation. Another factor relevant to the across-quantifier variation we observe in the current simulations is the assumption of cooperativity encoded in the model: listeners assume the speaker is being cooperative, which means that listeners assume that the speaker said something true. When interpreting an ambiguous utterance, then, listeners will be biased to resolve the ambiguity in a way that makes the utterance true. So, one source of interpretation preferences comes from preferring interpretations that will be true more often.

More specifically, the *not all* interpretations of *every*-negation (its inverse scope) and of *some*-negation (its surface scope) should be preferred to the alternative *none* interpretations. The reason is the same for both quantifiers: given the prior expectation that *all marbles are red*, both interpretations

of these quantifier-negation utterances convey that the prior high positive expectation is false (i.e., that $w \neq 3$). Learning that this strong prior belief is false is extremely informative. However, there are more ways for *not all* to be true (w could be 0, 1, or 2) than for *none* to be true (w must be 0). The listener reasons that the utterance is true, and so the speaker most likely intended the meaning that is true in more situations: the *not all* meaning (i.e., inverse for *every*-negation and surface for *some*-negation).

In contrast to the strongly-biased interpretation preferences predicted for *every*-negation and *some*-negation, *no*-negation is predicted to be more ambiguous, with no strong pressure toward either interpretation. Both its surface scope (*all*) and inverse scope (*some*) are compatible with the high positive expectation, and so are equally (un)informative because they convey that this expectation is true (and so the listener doesn't learn much). However, *all* is slightly preferred to *some* because *all* is most compatible with the high positive expectation (i.e., $w = 3$; *some* includes the possibility that $w = 1$ or 2, which are less compatible)—in this sense, *all* might then be viewed as (slightly) more likely to be true. For this reason, the *all* interpretation is then slightly preferred.

We note that another potential factor in interpretation differences across quantifiers is the particular logical relationship between the surface and inverse scope interpretations. For *every*-negation and *some*-negation, the model-predicted preferred interpretation is the entailed *not all* meaning. With *every*-negation, the surface scope interpretation (*none*) asymmetrically entails the inverse scope interpretation (*not all*): if *none* succeeded, then *not all* necessarily succeeded; in contrast, if *not all* succeeded, it might not be that *none* did. Likewise with *some*-negation, the inverse scope interpretation (*none*) asymmetrically entails the surface interpretation (*not all*), for the same reasons. However, entailment relations predict the opposite of the model for *no*-negation, because the surface scope *all* asymmetrically entails the inverse scope *some*: if *all* succeeded, then *some* necessarily succeeded; in contrast, if *some* succeeded, then it might not be that *all* did. So, entailment relations predict a preference for the inverse interpretation (*some*). This means that entailment relationships align with our model's predictions for two of three quantifiers.

Experiments

To test our modeled pragmatic listener's predictions about the preferred scope interpretation (which is captured by L_1's marginal posterior distribution over i), we measured interpretation behavior in a paraphrase-endorsement task similar to the corpus-annotation task but with invented utterances using the quantifiers *every*, *some*, and *no*. We first use a picture-selection task to validate the relevant paraphrases. We then asked participants to rate paraphrases corresponding to surface vs. inverse interpretations (e.g., *None/Not all of the marbles are red*) for a potentially-ambiguous utterance (e.g., *Every marble isn't red*). Figure 4 shows how the communication scenario was set up for both experiments.

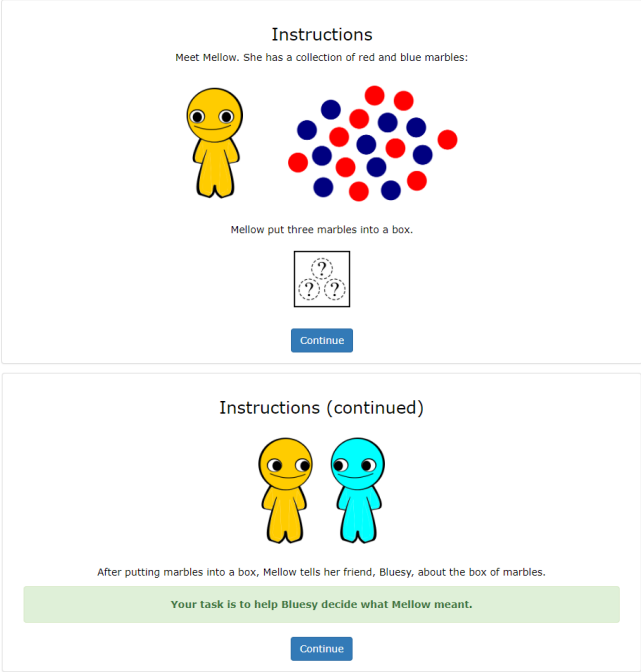

Figure 4: Communication scenario in experiments 1 and 2.

Experiment 1: Paraphrase validation

To verify that our paraphrases of scope interpretations for the paraphrase-endorsement task would be understood as the intended scope interpretation, we asked participants to complete a reference task (sample trial in Figure 5). Given a paraphrase, participants select the picture the paraphrase likely describes.

Participants. We recruited participants with U.S. IP addresses through MTurk. 95 participants (42% female; mean age: 37) indicated they understood the experiment and English was their only native language. Each received $0.50.

Materials. The surface/inverse paraphrases were these: *every: None/Not all of the marbles are red*; *some: Not all/None of the marbles are red*; *no: All/Some of the marbles are red.*

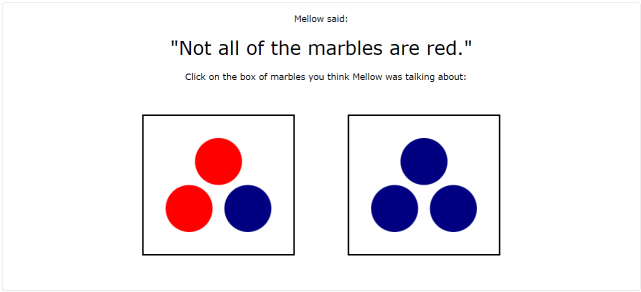

Figure 5: Reference task sample trial for inverse scope interpretation of *every*-negation in the paraphrase-validation task.

Design. The experiment began with a scenario intended to establish that the utterances to be interpreted were communication acts with a single likely meaning (Figure 4). Participants then saw an utterance and chose the scenario they thought the utterance described. For each utterance, participants chose between an image consistent with the surface scope and an image consistent with the inverse scope (e.g., between not-all-

red-marbles and no-red-marbles in Figure 5). The quantifiers *every*, *some*, and *no* were tested as a between-subject condition; participants completed a series of three trials in random order: one for the quantifier-negation utterance, one for the surface paraphrase, and one for the inverse paraphrase.

Results. Figure 6 shows the proportion of the time that participants chose the image indicating the inverse interpretation, grouped by utterance type (ambiguous, inverse, surface) and quantifier condition. Participants chose at ceiling the image consistent with the intended scope interpretation for each of the paraphrases (Figure 6: the middle panel shows inverse proportions near 1.0 for the inverse paraphrase and the right panel shows inverse proportions near 0.0 for the surface paraphrase). These results validate the paraphrases of *every*-negation, *some*-negation, and *no*-negation. For the potentially-ambiguous utterance, we found a non-significant trend (Figure 6, left panel) in line with the model predictions: *every* allows more inverse than *no*, which allows more inverse than *some*. We revisit this trend in the next experiment with a different and potentially more sensitive measure of interpretation preferences.

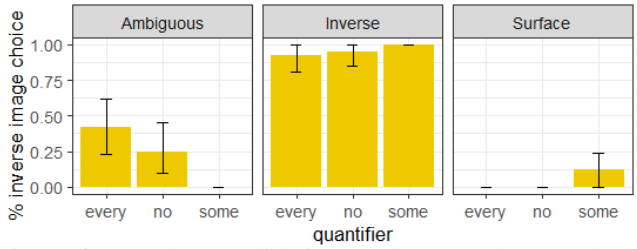

Figure 6: Paraphrase validation results. Error bars are bootstrapped 95% CIs.

Experiment 2: Paraphrase endorsement

We used the validated paraphrases to measure interpretation preferences for *every*-negation, *no*-negation, and *some*-negation utterances, following the paraphrase-endorsement methodology from Scontras and Goodman (2017).

Participants. We recruited 60 participants with U.S. IP addresses through MTurk. Of these, we assess data from the 47 participants (32% female; mean age: 36) who indicated they understood the experiment and English was their only native language. Each received $0.50.

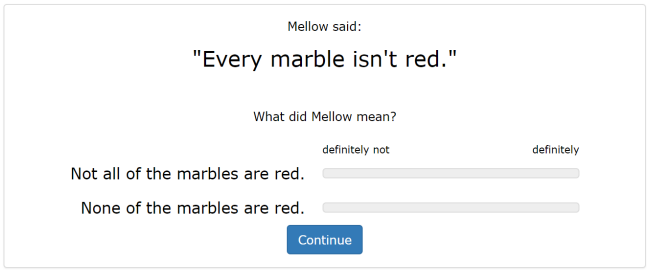

Figure 7: Sample paraphrase-endorsement trial.

Design. Participants saw the same communication scenario as in the previous experiment (Figure 4) and then rated two validated paraphrases of a potentially-ambiguous quantifier-negation utterance on a sliding scale (e.g., Figure 7), similar to the crowd-sourced corpus study. Participants completed three trials (one for *every*, *some*, and *no*) in random order.

Results. Figure 8 shows endorsement rates for validated surface and inverse paraphrases as well as model predictions for the marginal distribution over surface and inverse interpretations, grouped by quantifier. We found that ratings for the surface and inverse scope interpretations were negatively correlated per quantifier (*every*: -0.51; *no*: -0.40; *some*: -0.67), suggesting that endorsing one interpretation led to reduced endorsement for the other interpretation. To assess significance, we fit linear mixed effects models predicting the logit-transformed responses on each of the sliders by quantifier, with random intercepts for participant; all differences were significant. From left to right in Figure 8: *every* allowed the most inverse interpretations (95% CI [0.65, 0.84]), *no* allowed an intermediate proportion (95% CI [0.27, 0.47]), and *some* allowed the fewest (95% CI [0.07, 0.18]).

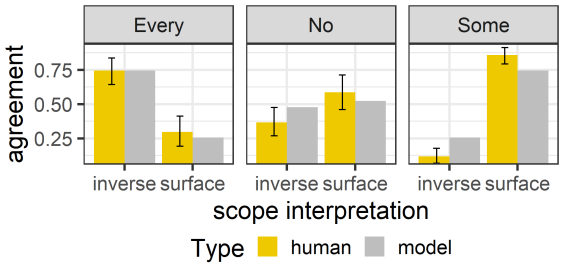

Figure 8: Results comparing model predictions and human data. Grey bars: Model predictions for L_1 marginal distribution over interpretation i, with $P(w = 3) = 0.9$. Yellow bars: Degree of endorsement of each paraphrase in the paraphrase-endorsement task. Error bars are bootstrapped 95% CIs.

We also see that the model predictions closely match the behavioral data. The model predicts 0.74 probability of inverse scope interpretations for *every*, 0.47 for *no*, and 0.26 for *some*. These predictions fall within the 95% CI for mean inverse scope probability for *every* and *no*, but slightly overpredict inverse scope interpretations for *some*. Model exploration revealed that the parameter setting allowing the model to capture the human data is the high positive expectation that all the marbles are red ($P(w = 3) = 0.9$).

Discussion. With a high positive expectation for the state of the world, the model predictions for the pragmatic listener's marginal distribution over scope interpretations align with interpretation patterns in the paraphrase-endorsement task: inverse scope is preferred for *every*-negation, slightly dispreferred for *no*-negation, and dispreferred for *some*-negation. These results support our model of ambiguity resolution and the importance of the high positive expectation. Importantly, our model not only captures the qualitative patterns in our data, but also largely captures the quantitative patterns.

General Discussion

We have strengthened the empirical basis for the idea of variation in quantifier-negation utterance interpretations. In par-

ticular, in a crowd-sourced corpus analysis of spontaneous speech, we find within-quantifier variation in interpretations of naturalistic *every*-negation utterances. Using behavioral experiments, we find across-quantifier variation in interpretations of quantifier-negation utterances with a universal (*every*), existential (*some*), and negative (*no*) quantifier.

We further find that a hypothesis of ambiguity resolution incorporating pragmatic factors, as formally articulated in our computational cognitive model, can explain the documented across-quantifier variation. Our model implements the hypothesis that interpretations (i) depend on a cooperative, efficient speaker, and (ii) build on certain prior expectations about the world. In particular, the model's ability to account for the data from our experiment depends on a high positive expectation about the world state (i.e., that all the marbles are red). With this expectation in place, the model predicts that the most likely interpretation is the most informative one that is most likely to be true: *not all* for *every*-negation, *none* for *some*-negation, and *all* for *no*-negation.

While our model used prior modeling work to set the precise value for the high positive expectation, the general idea that a high positive expectation is important for scope interpretation preferences finds support in our own *every*-negation corpus data. In particular, we found that the high positive expectation surfaced as content immediately preceding the potentially-ambiguous *every*-negation utterance; interestingly, this content often was the same linguistic form (e.g., *everything was OK*), so that the potentially-ambiguous utterance echoed a substantial part of that linguistic form (e.g., *everything wasn't OK*). Future work can continue to test the extent of repetition between the preceding discourse and ambiguous utterance, as well as explore patterns in the repeated content. To the extent that we can find positive expectations preceding quantifier-negation utterances, our findings will be in line with theories that negation use (such as the negation in quantifier-negation) is more felicitous in contexts that set up the corresponding affirmative information (e.g., Wason, 1961; Givón, 1978).

We also note that in the case of *some*, human responses were more categorical than the model predictions for *some*-negation, which may be due to *some*'s status as a positive polarity item (PPI) that doesn't scope under negation (Szabolcsi, 2004). However, our modeling results offer an explanation for why *some* might behave as a PPI in the first place: interpreting *some* under negation can result in an utterance that is uninformative, has an unlikely meaning, and is therefore inefficient.

Future work can continue expanding the empirical basis for variation in scope interpretations, exploring both within-quantifier and across-quantifier variation via the kind of crowd-sourced corpus analysis presented here. With computational modeling, we have shown how across-quantifier variation may be accounted for by the model's tendency to resolve ambiguity in a way that preserves truth; we hypothesize that within-quantifier variation of the sort documented in our preliminary corpus analysis arises through the interaction between contextual information and the model's tendency to resolve ambiguity in a way that yields more informative interpretations. Our findings thus underscore the usefulness of computational cognitive modeling for specifying the role that pragmatic factors, such as expectations about the state of the world, may play when we interpret ambiguous utterances in context.

Acknowledgments

We're grateful to anonymous CogSci reviewers and the QuantLang Collective at UCI for their helpful feedback on this work.

References

G. Carden. Disambiguation, favored readings, and variable rules. *New ways of analyzing variation in English*, pages 171–82, 1973.

A. Conroy, S. Fults, J. Musolino, and J. Lidz. Surface scope as a default: The effect of time in resolving quantifier scope ambiguity. In *Poster presented at the 21st CUNY Conference on Sentence Processing. University of North Carolina, Chapel Hill, March*, volume 13, 2008.

M. Davies. Corpus of Contemporary American English (COCA). 2015. URL `https://doi.org/10.7910/DVN/AMUDUW`.

J. Degen. Investigating the distribution of some (but not all) implicatures using corpora and web-based methods. *Semantics and Pragmatics*, 8:11–1, 2015.

M. C. Frank and N. D. Goodman. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998, 2012.

T. Givón. Negation in language: pragmatics, function, ontology. In *Pragmatics*, pages 69–112. Brill, 1978.

S. Lee. *Interpreting scope ambiguity in first and second language processing: Universal quantifiers and negation*. University of Hawaii at Manoa, 2009.

R. C. May. *The grammar of quantification.* PhD thesis, Massachusetts Institute of Technology, 1977.

J. Musolino. Universal grammar and the acquisition of semantic knowledge: An experimental investigation into the acquisition of quantifier-negation interaction in english. 1999.

J. Musolino and J. Lidz. Why children aren't universally successful with quantification. *Linguistics*, 44(4):817–852, 2006.

J. Musolino, S. Crain, and R. Thornton. Navigating negative quantificational space. *Linguistics*, 38(1):1–32, 2000.

K. Savinelli, G. Scontras, and L. Pearl. Modeling scope ambiguity resolution as pragmatic inference: Formalizing differences in child and adult behavior. In *CogSci*, 2017.

G. Scontras and N. D. Goodman. Resolving uncertainty in plural predication. *Cognition*, 168:294–311, 2017.

A. Szabolcsi. Positive polarity–negative polarity. *Natural Language & Linguistic Theory*, 22(2):409–452, 2004.

P. C. Wason. Response to affirmative and negative binary statements. *British Journal of Psychology*, 52(2):133–142, 1961.