Corpus evidence for the role of world knowledge in ambiguity reduction: Using high positive expectations to inform quantifier scope

Noa Attali, Lisa S. Pearl, & Gregory Scontras*

Abstract. *Every*-negation utterances (e.g., *Every vote doesn't count*) are ambiguous between a surface scope interpretation (e.g., *No vote counts*) and an inverse scope interpretation (e.g., *Not all votes count*). Investigations into the interpretation of these utterances have found variation: child and adult interpretations diverge (e.g., Musolino 1999) and adult interpretations of specific constructions show considerable disagreement (Carden 1973, Heringer 1970, Attali et al. 2021). Can we concretely identify factors to explain some of this variation and predict tendencies in individual interpretations? Here we show that a type of expectation about the world (which we call a *high positive expectation*), which can surface in the linguistic contexts of *every*-negation utterances, predicts experimental preferences for the inverse scope interpretation of different *every*-negation utterances. These findings suggest that (1) world knowledge, as set up in a linguistic context, helps to effectively reduce the ambiguity of potentially-ambiguous utterances for listeners, and (2) given that high positive expectations are a kind of affirmative context, negation use is felicitous in affirmative contexts (e.g., Wason 1961).

Keywords. scope ambiguity; universal quantifiers; negation; pragmatics; computational models; corpus linguistics; psycholinguistics

1. Introduction. It's unclear how people prefer to interpret ambiguous *every*-negation utterances, such as (1):

(1) Every vote doesn't count.

a. No vote counts. *Surface scope*: $\forall x[\text{vote}(x) \rightarrow \neg\text{count}(x)]$ (every > n't)

b. Not all the votes count. *Inverse scope*: $\neg\forall x[\text{vote}(x) \rightarrow \text{count}(x)]$ (n't > every)

The utterance in (1) allows both a surface interpretation (1a) and an inverse one (1b), depending on the logical scope of the quantifier relative to negation. Which interpretation would a listener or reader believe is more likely to be intended by the speaker?

Previous studies show variation in interpretation behavior. On the one hand, children and non-native speakers seem to disprefer the inverse scope interpretation of *every*-negation (Musolino 1999, Gualmini et al. 2008, Viau et al. 2010, Chung & Shin 2022); converging research on scope ambiguity suggests that surface scope is easier to access, involving less representational complexity or processing cost (Tunstall 1998, Pritchett & Whitman 1995, Anderson 2004, Lee et al. 2011). On the other hand, experimental studies that directly measure interpretation preference by adult native English speakers find a preference for inverse scope interpretations of *every*- and *all*-negation (Carden 1970, Heringer 1970, Carden 1973).

*We would very much like to thank the Quantitative Language Collective at UC Irvine for helpful comments and feedback on this work. Authors: Noa Attali, UC Irvine (nattali@uci.edu), Lisa S. Pearl, UC Irvine (lpearl@uci.edu) & Gregory Scontras, UC Irvine (g.scontras@uci.edu).

Additionally, the above studies report preferences that are aggregated across different sentences and contexts. When we look beyond average interpretations to behavior on individual sentences, we find further variation. Changes in the immediate linguistic context – in fact, in the sentence containing the quantifier-negation construction itself – can flip interpretation patterns. Carden (1973) found that for (2), 82.5% of respondents said that only the inverse scope interpretation was possible, 7.5% said that both were possible but that they favored inverse scope, and none reported accessing only surface scope. On the other hand, for (3), 100% said that only the surface scope interpretation was possible.

(2) All the boys didn't arrive, did they?

(3) All the boys didn't leave until midnight.

This variation highlights how findings on interpretation patterns depend on the stimuli themselves. What then are the characteristics of the contexts that impact these interpretation preferences for *every*-negation utterances? That is, when are particular scope interpretations preferred?

In their computational cognitive model of scope ambiguity, Scontras & Pearl (2021) demonstrate that a kind of world knowledge we term a "high positive expectation" (**HPE**) can explain some variation in interpretation preferences with *every*-negation utterances. To illustrate, an HPE for *Every vote doesn't count* is the prior belief that it's highly likely that every vote *does* count. That is, the HPE for this *every*-negation utterance is that the worlds consistent with the equivalent positive utterance (*Every vote does count*) have a high probability. In general, an HPE is the prior belief that the most likely true world states are those consistent with the *every* world state in the universe of the utterance.

An HPE could contribute to the felicity of using *every*-negation with an inverse scope *not all* interpretation (e.g., Not all the votes count), thereby reducing the ambiguity of the utterance for listeners. This ambiguity reduction occurs because the *not all* interpretation is highly informative about the expectation that *all* is true (by indicating that this expectation isn't correct). We might put it this way: in a context with this kind of expectation (e.g., All votes count), the inverse scope interpretation of *every*-negation (e.g., Not all votes count) is felicitous as an emphatic message that the salient expectation (e.g., that all votes do in fact count) is false. As Scontras & Pearl (2021) suggest, this world knowledge factor is one way to quantitatively specify a pragmatic factor that explains prior behavioral results (see Scontras & Pearl 2021 for more details).

Scontras & Pearl's model implements a cooperative, efficient speaker – cooperative in wanting to say something true, and efficient in wanting to be informative. For example, returning to the counting votes case, the model predicts that when the context holds the expectation that all votes count, speakers would more likely agree that *Every vote doesn't count* could be used to mean *Not all votes count*. Given this HPE context, the inverse scope interpretation of *Every vote doesn't count* felicitously – i.e., cooperatively and efficiently – conveys that some votes do not, in fact, count. In general, the model predicts that speakers tend to endorse *every*-negation (e.g., *Every vote doesn't count*) as a true description of a scenario consistent with the inverse scope interpretation (e.g., Not all votes count) when *every*-negation conveys that an HPE (e.g., All votes count) is false (e.g., It's false that all votes count).

The model's predictions for the role of an HPE extend well to prior findings. Attali et al. (2021)

find that Scontras & Pearl's model successfully predicts listener interpretations of *every*-negation when given an HPE. Specifically, the modeled listener shows a close qualitative and quantitative match to the average preference for the inverse scope interpretation of an out-of-context sentence *Every marble isn't red*. The inverse scope *not all* interpretation is preferred over surface scope *none* when given an HPE because (1) there are more ways for the *not all* interpretation to be true compared with the *none* interpretation, and so a cooperative speaker intends that interpretation, and (2) it's highly informative to update a strongly biased, salient belief such as an HPE (e.g., that every vote does count), and so an efficient speaker intends that interpretation; for more details see Attali et al. 2021).

Because HPEs can explain interpretations for *every*-negation utterances in prior experimental and computational work, we ask how well HPEs account for interpretations of *every*-negation utterances in naturalistic contexts. For instance, do HPEs surface in the contexts of spontaneously produced *every*-negation? Specifically, when a local linguistic context seems to express an HPE, is an inverse scope interpretation more likely than a surface scope interpretation?

We first describe how we developed a corpus of naturally-occurring *every*-negation uses in context via a behavioral study, including the preferred interpretation of each use. We then describe how we identified HPEs in the corpus contexts. We present our analysis for the connection between the presence of an HPE and an item's preferred interpretation, finding that inverse scope interpretations are indeed more preferred following an HPE. Our results suggest that the world knowledge implemented as an HPE in the local linguistic context can help effectively reduce the ambiguity of *every*-negation utterances for listeners, and that negation use is felicitous in affirmative contexts, like the kind that HPEs encode.

2. Corpus data and behavioral experiment. To assess the role of HPEs in the interpretation of naturally-occurring *every*-negation utterances, we need (1) a corpus of naturally-occurring *every*-negation utterances in context, (2) a measure of these utterances' preferred interpretations, and (3) an estimate of the extent to which a context contains an HPE. In this section, we describe how we created a corpus (goal 1) and measured interpretation preferences in context (goal 2). Section 3 discusses how we identified HPEs in context (goal 3) to see whether their presence predicts inverse scope interpretations.

2.1. CORPUS SEARCH FOR *every*-NEGATION UTTERANCES. To achieve goal 1, we identified uses of *every*-negation in a corpus of spontaneous speech. We extracted the *every*-negation occurrences in the speech section of the Corpus of Contemporary American English (COCA; Davies 2015), which is made up of transcripts of spoken conversations from American radio and TV programs from 1990 to 2012 (≈9 million clauses, or ≈95 million words). We defined *every*-negation occurrences as those where quantified subjects precede and c-command sentential negation (with *not* or contracted *n't*). To develop the automated search, we randomly selected a year of COCA transcripts and manually searched it for uses of *every*-negation. We then wrote a search that yielded a 100% recall rate, that is, returning each of the occurrences in this development set. We applied this search to the rest of the COCA speech section. We found that *every*-negation uses are highly infrequent in English conversation; in total, we identified 390 cases.

2.2. CORPUS ANNOTATION. To achieve goal 2, following Degen (2015), we crowd-sourced interpretation preferences of these uses in their immediate contexts (three preceding sentences and one following sentence). For each item, participants (N = 208) completed a paraphrase-endorsement task (Scontras & Goodman 2017), choosing on a sliding scale between *none* and *not all* paraphrases of the potentially-ambiguous clause.

2.2.1. PARTICIPANTS. We recruited 390 participants with U.S. IP addresses through Amazon.com's Mechanical Turk (MTurk) crowd-sourcing service. Of these 390, we assessed data from 208 participants (35% female; mean age: 41 y/o) who passed control trials (discussed further in Section 2.2.4) and indicated that English was their only native language. Each received $2.00.

2.2.2. STIMULI. For each of the 390 *every*-negation uses, we created excerpts consisting of the three preceding context sentences, the bolded potentially-ambiguous clause, and one following context sentence (see Figure 1). We also created paraphrases of the surface and inverse scope interpretations. As part of a pilot experiment (N=94), we checked that the paraphrase wording was correctly understood, so that the surface scope interpretation paraphrase was always understood to be consistent with a *none* situation (e.g., that *None are red* describes three blue marbles rather than two red and one blue marble) and the inverse scope paraphrase was always understood to be consistent with a *some but not all* situation (e.g., that *Not all are red* preferentially describes two red and one blue marble rather than three blue marbles).

Because the ambiguous clauses took the form *quantified noun phrase–verb–negation–remainder* (e.g., *Everybody is not doing that*: quantified noun phrase = *Everybody*; verb = *is*; negation = *not*; remainder = *doing that*), surface scope paraphrases took the form *none/no one/nobody/nothing–verb–remainder* (e.g., *nobody is doing that*) and inverse scope paraphrases took the form *not all/not all things are–remainder* (e.g., *not all are doing that*).

Figure 1: Sample paraphrase-endorsement trial from the corpus analysis of *every*-negation.

2.2.3. DESIGN. The initial instructions asked to "choose the best paraphrase for the bolded part" for fifteen randomly-selected conversation excerpts; at each trial, participants were again asked "What did the speaker mean in the bolded part?" (see Figure 1). Beneath the excerpt, participants rated the best paraphrase as a judgment on a sliding scale between the surface and inverse scope interpretations. The two scope interpretations were randomly assigned for each item in left-right or right-left order.

2.2.4. CONTROLS. To check that participants were reading and understanding the contexts of the items – and also as a way to suggest that context is useful – two control trials were constructed to imitate the items from the corpus. These control trials contained clearly disambiguating information about the intended scope interpretation in the surrounding context. The disambiguating information always appeared as a restatement of the speaker's meaning. These two controls appeared in random order as the first two trials for each participant.

The surface scope-disambiguating control item is in (4), and the inverse scope-disambiguating control item is in (5). For clarity, the disambiguating information is italicized, though it was not italicized in the experiment. Participants were considered to pass the surface control by placing the slider closer to the *none* paraphrase than to the *not all* paraphrase; they passed the inverse control by placing the slider closer to the *not all* paraphrase than to the *nobody* paraphrase.

(4) TONHAUSER: The ten board members voted last night. I was really surprised—I thought at least some of them would like Proposition 23. But *all ten of them voted against it*. Basically, **every board member didn't like Proposition 23**. *Not even a single one of them liked it*.

(5) SIDNER: Look, we completely fixed the issue. Indicators have improved across the board. Everybody's happy.
GROSZ: (VOICEOVER) No, **everybody isn't happy**. *Some are happy but others are deeply dissatisfied with what they call a 'band aid solution.'*

We restricted analysis to those participants who passed both controls and indicated that English was their only native language. The rate of passing both controls was 53%. This relatively high failure rate may have been due to low English reading proficiency. Though we restricted MTurk participation to US IP addresses and to those MTurk workers who have completed at least 1,000 tasks in the past and we also only analyzed data from self-reported native English speakers, it's possible that some participants didn't fluently read English well enough. Another factor may have been attention and motivation: participants in an online platform may be disengaged with the experiment. A third factor is task difficulty: perhaps the paraphrase endorsement task could be seen as a complex reading comprehension and logical inference task, because these sentences have multiple logical operators.

2.3. RESULTS. Each item was judged by at least 2 and at most 14 different participants. Although the surface scope paraphrases randomly appeared on the left or right of the sliders, we transformed and report responses on sliders as though the surface scope paraphrases always appeared on the left. This allows the final response measure to vary from 0 (maximum endorsement of the surface scope interpretation) to 1 (maximum endorsement of the inverse scope interpretation).

In the COCA transcripts, we found both a general preference for inverse scope interpretations as well as a high degree of interpretation variation for the *every*-negation utterances. Figure 2, which shows judgment-by-judgment interpretations, suggests that many of these utterances in context seem unambiguous: 29% of individual scores were below 0.25 (indicating a strongly surface scope interpretation) while 53% of individual scores were above 0.75 (indicating a strongly inverse scope interpretation). Figure 3 aggregates ratings by items. For some items, strong intuitions are reliable across different participants' judgments: 12% of mean scores were below 0.25, and 38% of mean scores were above 0.75. Examples (6) and (7) are items that elicited a strong

surface scope preference ((6): mean response ≈ 0.13) or a strong inverse scope preference ((7): mean response ≈ 0.98).

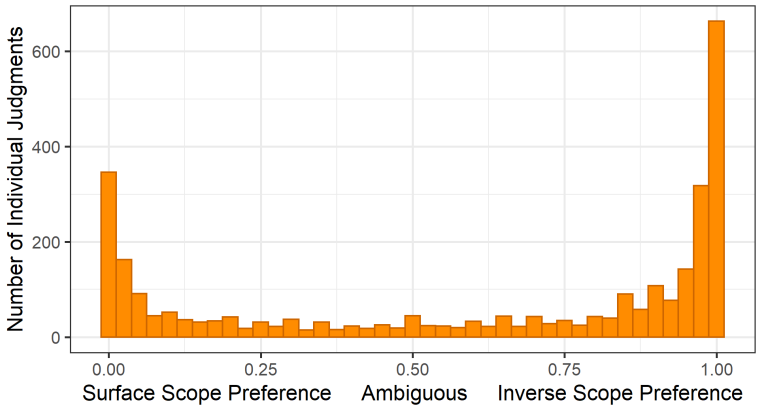

Figure 2: Individual scope interpretations from the *every*-negation corpus analysis.

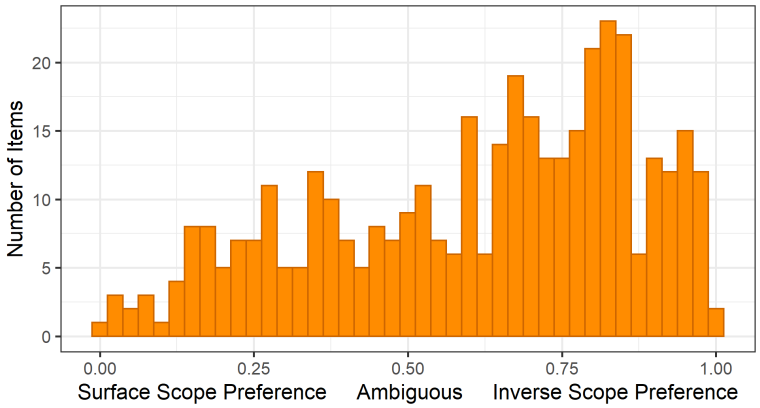

Figure 3: Mean interpretations per item from the *every*-negation corpus analysis.

(6) @!WERTHEIMER: So what about New Jersey? Can New Jersey get over the hump?
@!RAPOPORT: Well, [transcript cuts out] first team from the Eastern Division to return to the finals since the Bulls were winning all their championships. They're a little nervous about that in New Jersey, Linda, that **every team that made it to the finals from the East in the last couple of years hasn't been able to repeat**; but again, they're strong competition. The Pistons have been impressive this year.

a. No team (that made it to the finals from the East in the last couple of years) has been able to repeat. *(every > n't)*

b. Not all teams (that made it to the finals from the East in the last couple of years) have been able to repeat. *(n't > every)*

(7) @!CALLER Hi. My question for Mr. Eisner was, MGM is one of my favorite places in Disneyworld and one of my favorite attractions there is the animation studios, and now the studio, the animation studio there is closed, and everything has moved to California, and I

wanted to know how you justified doing that.
@!EISNER Well, **everything has not moved to California**. We will still be demonstrating animation in Florida.

a. Nothing has moved to California. *Surface scope* (every > not)
b. Not all things have moved to California. *Inverse scope* (not > every)

2.4. DISCUSSION. The results of the paraphrase endorsement study with the corpus-mined stimuli show variation and a weak inverse scope preference in adult native English speakers' interpretations for *every*-negation utterances. Although these results agree with the larger picture painted by previous studies on *every*-negation, to our knowledge this is the first larger-scale investigation of naturalistic stimuli in context.

In the following section, we ask whether HPEs account for this variation and inverse scope interpretation preference. More specifically, when an inverse scope interpretation is preferred for an *every*-negation utterance (e.g., Not all votes count for *Every vote didn't count*), was that use of *every*-negation in fact an emphatic message that a salient HPE (e.g., All votes count) is false?

3. Identifying high positive expectations in linguistic contexts. An HPE represents a prior belief, and one way to measure for its presence is by its overt linguistic expression in an item's preceding context. For example, for *Every vote doesn't count*, an HPE is the prior belief that every vote *does* count – that is, that the worlds consistent with the non-negated utterance (*Every vote does count*) are highly probable – and one measure of this HPE's presence is the non-negated utterance itself: *Every vote does count*.

As a preliminary measure, the first author hand-coded categorically for the presence/absence of an overt HPE expression in each preceding context. We found that 59/390 (15%) of the items contained such an expression.

For an automatic and more objective measure of the HPE expression – that is, a method that could scale to large amounts of data and would capture the intended linguistic phenomenon while minimizing experimenter bias – we calculated the degree of lexical overlap between the preceding linguistic context and a string representing the positive expectation (pos_exp). For each item (e.g., *Every vote doesn't count*), we first coded pos_exp as the potentially-ambiguous clause without negation (e.g., *Every vote does count*). We then coded for the extent to which the pos_exp appeared in the preceding context as the longest common substring similarity (**LCS**; Needleman & Wunsch 1970) between each preceding context string c and pos_exp pair, calculated using the R `stringdist` package (Loo 2014).

Each LCS was equal to the longest sequence formed by pairing words from the preceding context string c and pos_exp, while keeping their order intact; the dissimilarity $d_{lcs}(c, pos\_exp)$ was then the number of unpaired words left over in both strings. $d_{lcs}(c, pos\_exp)$ can be defined recursively as in (8) for different relative lengths of the two strings to be matched against:

(i) It is trivially 0 for empty strings (line 1: $c = pos\_exp = \epsilon$).

(ii) It is based on pairing each word from both strings if the two strings have equal length (line 2: $|c| = |pos\_exp|$). For example, suppose the preceding context is *Every vote does count* for an utterance with the pos_exp *Every vote does count*. Then, d_{lcs}= 0. However, if the preceding context was *What is going on?*, d_{lcs}= -8 because all eight words in the two strings would be unpaired.

(iii) It is based on the minimum lcs-distance that can be obtained from pairing all the words

from the shorter string to an equal number of words from the longer string (line 3: otherwise). For example, suppose the preceding context was *I believe every vote does count* for an utterance with the pos_exp *Every vote does count*. Then, d_{lcs}= -2 because all four words in pos_exp would pair to *every vote does count* in the context, and leave unpaired the two words *I believe*.

Thus, dissimilarity ranges from 0 (completely similar) to the total words W in both strings combined (completely dissimilar), where $W = (|c| + |pos\_exp|)$. We calculate LCS similarity as negative dissimilarity: $-d_{lcs}(c, pos\_exp)$. Thus, LCS similarity ranges from 0 to -W, with values closer to zero indicating more lexical overlap. In particular, values closer to zero indicate a greater similarity between the context and the HPE linguistic string, and so represent a higher probability that the context contained a linguistic string transparently encoding an HPE.

$$d_{lcs}(c, pos\_exp) = \begin{cases} 0, & \text{if } c = pos\_exp = \varepsilon \\ d_{lcs}(c_{1:|c|-1}, pos\_exp_{1:|pos\_exp|-1}), & \text{if } |c| = |pos\_exp| \\ 1 + min\{d_{lcs}(c_{1:|c|-1}, pos\_exp), & \\ \qquad d_{lcs}(c, pos\_exp_{1:|pos\_exp|-1})\}, & \text{otherwise.} \end{cases} \tag{8}$$

4. Results.

4.1. HAND-CODED HPE RESULTS. Using the preliminary categorical hand-coding where we found that 59/390 of the utterances had HPEs, we first looked at p(inverse|HPE): how often an inverse scope interpretation was preferred when an HPE occurred. We found that 50/59 (85%) of utterances with HPEs were on average better paraphrased by the inverse scope paraphrase *not all* than the surface scope paraphrase *none*.

We also looked at p(HPE|inverse) vs. p(HPE|surface): how often items where the inverse interpretation was strongly preferred had an HPE compared with items where the surface interpretation was strongly preferred. We found that 22% of highly inverse-preferred items (those with responses greater than 0.75) had HPEs, as opposed to 6% of highly surface scope-preferred items (those with responses less than 0.25).

These results suggest that the hand-coded HPEs do tend to co-occur with an inverse scope interpretation in our sample. However, the automatic measure of an HPE's presence that we described above allows us to to measure the continuous relationship between the extent of HPE expression and the strength of inverse scope preference, as shown below. In addition, this measure can be used in future work to analyze larger samples.

4.2. AUTOMATIC HPE RESULTS. We used the continuous LCS-based measure $-d_{lcs}$ to assess if an HPE predicts an inverse scope preference per item, and ran a linear mixed effects model predicting logit-transformed mean item responses by $-d_{lcs}$ (representing LCS similarity) with random intercepts for participants (see Figure 4). To determine whether an HPE captures individual judgment variation above and beyond mean item-level variation, we predicted logit-transformed item responses by LCS similarity, with random intercepts for participants and items. Both models found that LCS similarity was a significant predictor of an inverse scope preference preference ($p < .001$ in both). That being said, although LCS similarity is a significant predictor of inverse scope preference, the relationship is noisy, as Figure 4 shows, with a marginal $R^2 = 0.024$.

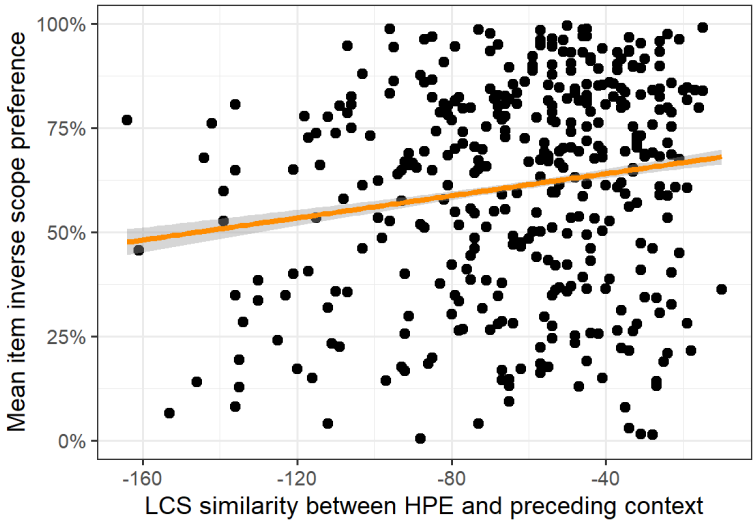

Figure 4: Preceding HPE and average inverse scope item preference.

Interestingly, only preceding, and not following, expressions of HPEs predict an inverse scope preference, as Figure 5 shows: a version of both models that calculated LCS similarity using overlap with the following – rather than preceding – context, found LCS similarity of the following context not to be a significant predictor of either item-level or judgment-level interpretations.

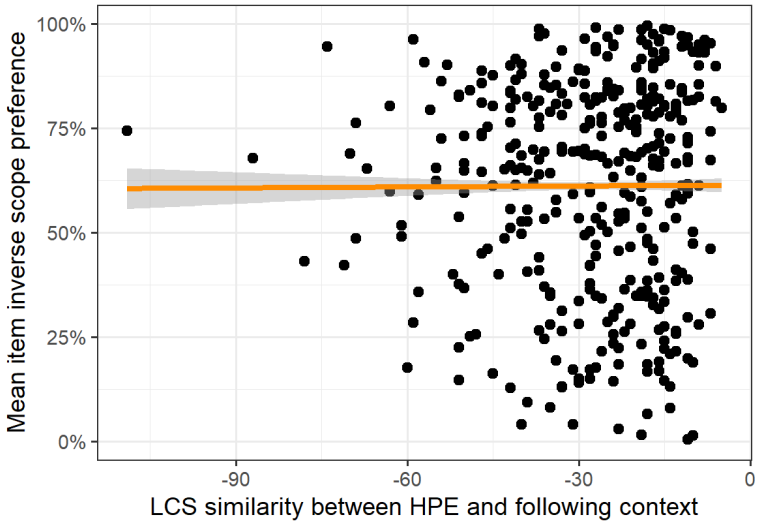

Figure 5: Following HPE and average inverse scope item preference.

5. Discussion. Our corpus analysis suggests that a high positive expectation (**HPE**) expressed directly in the preceding linguistic context can affect scope interpretation preferences for *every*-negation utterances. In particular, HPEs expressed this way correlate with stronger preferences for the inverse interpretation. These results align with previous modeling results (Scontras & Pearl 2021) and pragmatically-oriented proposals from truth-value judgment studies for supporting the

felicity of *every*-negation in context (Gualmini et al. 2008). In particular, an HPE provides a context that makes an inverse scope interpretation more felicitous because several things hold: (1) the listener assumes that speakers are truthful and informative, (2) an HPE represents a strong belief about the world, (3) finding out an HPE is false (that *not all* is true) is very informative, and (4) the inverse scope *not all* interpretation is one way to find this out. We speculate that perhaps in comparison with alternative constructions such as *Not every* (e.g., *Not every vote counts* for *Every vote doesn't count*), the *every*-negation construction highlights that a positive expectation is false, and might even be preferred as more informative (in such a context) than its *not every* alternative.

We note that our LCS similarity measure for HPEs is a first-pass one (the first anyone has tried to our knowledge), and likely underestimates HPE presence. In particular, this measure looks for transparent linguistic encodings of an HPE; but of course world expectations do not have to be encoded linguistically, or encoded nearby even if they are linguistically encoded. Even given our restriction to overtly expressed world knowledge in the preceding three sentences, our specific measurement of LCS similarity has a noisy potential to underestimate the presence of an HPE for several reasons. First, it is affected by context length, such that LCS similarity is lower for longer contexts even if those contexts contain a clear HPE expression. For instance, returning to the vote-counting case, LCS similarity would be -2 for the context *I believe every vote does count* but it would be 0 the context *Every vote does count*. Second, this LCS similarity measure looks for an HPE based on the exact lexical items in the *every*-negation utterance. For instance, it would identify the HPE in the context *Every vote does count* for the *every*-negation utterance *Every vote doesn't count*; yet, this measure misses the HPE in the context *All votes should matter* because the individual lexical items differ (*every* vs. *all*, *count* vs. *matter*). This rigidity of LCS similarity as a measure of context-sentence overlap could be a source of the noisiness evidenced in Figure 4. Other potential sources of noise include additional factors that may help disambiguate scope interpretations, such as prior expectations about questions under discussion and grammatical scope accessibility (e.g., as found by Scontras & Pearl 2021).

Still, the advantage of LCS similarity is that it provides an automatic continuous measure to improve our analysis of larger-scale data. Here, it allowed us to consider the potential linear relationship between the extent of HPE expression and the extent of an inverse scope preference. Future work could replace LCS similarity with a measure that considers a vectorized *semantic* representation of meaning rather than lexical overlap between the context and a string representing the HPE. A vectorized semantic measure would allow for the flexibility to recognize degrees of semantic similarity rather than categorical lexical equivalence. For example, such an approach would allow us to count *All votes should matter* as a context expressing an HPE for *Every vote doesn't count* (recognizing that *all* is similar to *every*, and *count* similar to *matter*, in this context).

More generally, our results suggest that listeners can rely on world knowledge and properties of the immediately surrounding contexts of an ambiguous utterance, like an *every*-negation utterance, to interpret it. Since a context containing an HPE is a kind of affirmative context, these findings support the broader theory that negation use is more felicitous in affirmative contexts (e.g., Wason 1961). One way that listeners understand *every*-negation in context is as a kind of emphatic frame for the message that an HPE is false.

References

Anderson, Catherine. 2004. *The structure and real-time comprehension of quantifier scope ambiguity*: Northwestern University Evanston, IL dissertation.

Attali, Noa, Gregory Scontras & Lisa S Pearl. 2021. Pragmatic factors can explain variation in interpretation preferences for quantifier-negation utterances: A computational approach. *Proceedings of the Annual Meeting of the Cognitive Science Society* 43(43).

Carden, Guy. 1970. A note on conflicting idiolects. *Linguistic Inquiry* 1(3). 281–290.

Carden, Guy. 1973. Disambiguation, favored readings, and variable rules. *New Ways of Analyzing Variation in English* 171–82.

Chung, Eun Seon & Jeong-Ah Shin. 2022. Native and second language processing of quantifier scope ambiguity. *Second Language Research* 1–26.

Davies, Mark. 2015. Corpus of Contemporary American English (COCA) https://doi.org/10.7910/DVN/AMUDUW.

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

Gualmini, Andrea, Sarah Hulsey, Valentine Hacquard & Danny Fox. 2008. The question–answer requirement for scope assignment. *Natural Language Semantics* 16(3). 205.

Heringer, James T. 1970. Research on quantifier-negative idiolects. *Chicago Linguistic Society* 6. 95.

Lee, Miseon, Hye-Young Kwak, Sunyoung Lee & William O'Grady. 2011. Processing, pragmatics, and scope in korean and english. *Japanese/Korean Linguistics* 19. 297–311.

Loo, Mark P.J. van der. 2014. The stringdist Package for Approximate String Matching. *The R Journal* 6(1). 111–122. 10.32614/RJ-2014-011. https://doi.org/10.32614/RJ-2014-011.

Musolino, Julien. 1999. *Universal grammar and the acquisition of semantic knowledge: An experimental investigation into the acquisition of quantifier-negation interaction in english*: University of Maryland, College Park dissertation.

Needleman, Saul B & Christian D Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* 48(3). 443–453.

Pritchett, Bradley & John Whitman. 1995. Syntactic representation and interpretive preference. *Japanese Sentence Processing* 65–76.

Scontras, Gregory & Noah D Goodman. 2017. Resolving uncertainty in plural predication. *Cognition* 168. 294–311.

Scontras, Gregory & Lisa S Pearl. 2021. When pragmatics matters more for truth-value judgments: An investigation of quantifier scope ambiguity. *Glossa: a journal of general linguistics* 6(1).

Tunstall, Susanne Lynn. 1998. *The interpretation of quantifiers: semantics & processing*: University of Massachusetts at Amherst dissertation.

Viau, Joshua, Jeffrey Lidz & Julien Musolino. 2010. Priming of abstract logical representations in 4-year-olds. *Language Acquisition* 17(1-2). 26–50.

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