Speaker Knowledge Influences the Comprehension of Pragmatic Inferences

Replication of Bergen & Grodner 2012

Disha Dasgupta

Table of Contents

Introduction/Background

Conversations require the perceiver to extract both explicit and implicit meanings from a speaker's statement. While the explicit meaning comes from what the speaker literally says, the perceiver must derive the implicit meaning. In deriving this implicit meaning, the perceiver must incorporate their perception of what the speaker knows. I exemplify this with the following statement:

(1) Some of the dancers were tired.

Beyond the literal meaning of the statement, the perceiver draws an inference about why the speaker chose to use the word "some" instead of the strong statement "all" – these inferences are known as *scalar implicatures*. There are two types of scalar implicatures that the perceiver can produce in which the perceiver assumes the speaker *knows for sure* not all of the dancers were tired (strong implicature) or the speaker is *unsure* whether all the dancers were tired (weak implicature).

But how does this perceived speaker knowledge level affect the generation and comprehension of scalar implicatures? In 2012, Leon Bergen and Daniel J. Grodner conducted a study investigating this very question[1]. And through my replication, I researched the same question, aiming to recreate the same results Bergen & Grodner observed in the original study[2]. To answer this overarching question, the researchers and I posed the following hypothesis:

"If perceivers incorporate speaker knowledge during online comprehension, then they will generate a strong implicature in the full-knowledge context [speaker knows for sure] but a weak implicature in the partial-knowledge context [speaker is unsure]."

To test this hypothesis, the researchers and I proposed a set of linking assumptions designed to associate reading times (RTs) over stimulus passages with knowledge contexts and implicature generation. An example stimulus passage and regions of interest relevant to the linking assumptions are provided in Table 1.

Our first linking assumption poses that in the full-knowledge context, RTs over the scalar quantifier should be higher than in the partial knowledge context. In the full-knowledge context, the scalar quantifier should convey more information; specifically, the scalar quantifier should convey that *some of and not all…* where in the partial-knowledge context, the scalar quantifier should only convey that *some of*. The second linking assumption proposes an observed decrease in RTs over the complement sentence in the full-knowledge context. This ties back to the *some of and not all* condition the scalar quantifier conveys in the full-knowledge condition. Because of the *not all* implication, it is assumed that the strong implicature evokes the complement set (*the rest of…*) during the trigger sentence. The weak implicature does not, so the complement set will be generated and processed while reading the continuation sentence in the partial-knowledge context. With our final linking assumption, we assert that in the focused trigger type condition, trigger and complement sentence processing will remain unaffected by speaker knowledge. This is because the focused trigger asserts the *not all* condition in a literal manner. This literal statement indicates that regardless of speaker knowledge, the speaker knows for sure the *not all* condition, and subsequently, the complement set will also be evoked in both knowledge contexts.

Sentence type	Example
Context	
Full-knowledge	At my client's request, I meticulously compiled the investment report.
Partial-knowledge	At my client's request, I skimmed the investment report.
Trigger	
Scalar	Some of the real_estate investments lost money.
Focused	Only some of the real_estate investments lost money.
Continuation	
Complement	The rest were successful despite the recent economic downturn.

Table 1. A sample stimulus passage from the experiment, consisting of three sentences: the context, trigger, and continuation sentence. The context sentence provides information regarding the speaker's knowledge level, as indicated by the difference between *meticulously compiled* and *skimmed*. The trigger sentence stipulates the *not all* implication from the speaker's statement. Both the scalar and focused trigger sentence types include *quantifier* ("S/some of") and *spillover* ("the real estate") regions. The focused trigger sentence type includes an extra *focus particle* ("Only") explicitly stating the *not all* implication. The stimulus passage ends with a continuation sentence, which includes an *anaphor* ("The rest") and a subsequent *predicate* ("were successful") region.

[2] Links to the GitHub project repository and corresponding preregistration

Methods

Participants

In the original study, Bergen & Grodner collected data from 42 native English-speaking students from Swarthmore College, paying each student $8. For my replication, I collected data from 52 native English-speaking adults, paying each participant $6. While the 10 extra trials were initially run to account for potentially incorrectly completed trials, none of the data was compromised and I used the data from all 52 participants. One discrepancy here is that the original researchers also chose to run 20 extra participants to help resolve the extent of speaker knowledge in the partial-knowledge context; I chose not to run this subset of participants for time concerns.

Materials

The experiment consisted of 24 stimulus passage trials (each with 3 sentences: the knowledge context, trigger, and continuation sentence) as well as 65 additional filler trials, all of which were provided in the original study. Within the experiment, each of the 24 stimulus passages followed one of 6 conditions. The first 4 conditions were generated by crossing the full/partial knowledge context sentences and scalar/focused trigger sentences. I generated the final 2 conditions by duplicating the conditions with the scalar trigger sentences, then replacing the complement continuation sentences with cancelation continuation sentences. These cancelation sentences were designed to contradict the *not all* implication from the previous trigger sentence.

Procedure

Each of the 24 stimuli passages were randomly assigned to one of the 6 conditions in a Latin square design such that no participant saw any stimulus passage more than once or under a different condition[3]. All the 89 trials were then shown in randomized order to each participant. Every trial followed the self-paced reading paradigm, for which the implementation setup was generously provided by Sebastian Schuster. Given this paradigm, participants were shown one

[3] While I implemented the Latin square design, I did not account for showing stimuli in each of the 6 conditions equally. This imbalance may have contributed to potential discrepancies mentioned in the subsequent *Results* section.

word at a time, while I recorded the RTs for each word, (indicated by the participant pressing the space bar to progress to the next word). Following each trial, participants were asked a corresponding comprehension question. Finally, at the end of the study, I collected demographic information for each participant, primarily to reaffirm they were native English speakers but also for potential exploratory purposes.

While the data collection methodology was the same as in the original study, is important to note differences between my subsequent analyses and the analyses conducted by Bergen & Grodner. The original study first included a comparative analysis of the mean RTs at each of the specific regions of interest in the trigger, complement, and cancelation sentences. To this extent, I chose to replicate the comparative analyses for the trigger and complement sentences, as these are the sentences particularly highlighted within the original paper. The original researchers also conducted a series of linear mixed-effect regression analyses predicting log-transformed RTs from speaker knowledge and trigger type. These analyses modeled speaker knowledge and trigger type as fixed effects and participants and stimulus items as random effects. The primary difference in analysis was that the original researchers modeled *each* of the sentence regions, not only the previously highlighted regions of interest in Table 1, for the trigger, complement, and cancelation sentences. Again, in the interest of time, I chose to implement a subset of these analyses, running linear mixed-effect regression models on the specified regions of interest for the trigger and complement sentences.

Results

As a preliminary analysis, the researchers calculated the percentage of comprehension questions participants answered correctly. They found 91% of participants did so, and in my replication, I found 88% of participants did so. Although not exact, slight variability is expected, and thus I believe this finding was replicated.

Comparative Analysis of Mean RTs

Our first two linking assumptions hypothesized increased RTs over the scalar quantifier in the trigger sentence and decreased RTs over the complement sentence in the full-knowledge context. Through my replication, we can observe how the reading times compared in these regions of interest in the full and partial-knowledge contexts. From Figure 1, we observed that in the scalar trigger type condition, my results generally follow the original results, specifically also that the quantifier region was read faster in the partial-knowledge condition than in the full-knowledge condition. As indicated by lack of overlap of the error bars (95% CI) in both the original study and my replication, the observed increase in RTs in the full-knowledge condition is significant.

In contrast, as indicated in Figure 2, I noticed that within the scalar trigger type condition, every region in the complement sentence was read faster in the partial-knowledge condition than in the full-knowledge condition, contradicting the original researchers' findings as well as the second proposed linking hypothesis. Again, as indicated by the nonoverlapping error bars, this contrasting difference in RTs over the complement sentence I found is significant.

The final linking assumption Bergen & Grodner and I proposed is that in the focused trigger type condition, speaker knowledge would not affect RTs over the focus particle and quantifier regions in the trigger sentence and the anaphor in the complement sentence. From my corresponding results depicted in Figures 3 and 4, we see two primary findings. First, the qualitative patterns of my results generally corresponded with those in the original study, excluding the final region of the complement sentence in the partial knowledge condition. Second, as indicated by the large overlap in the error bars between the knowledge contexts, there were no significant differences in RTs over the focus particle and quantifier regions in the trigger sentence and the anaphor in the complement sentence. This result indicates a general replication of the original results, but not to the full extent given some of the deviations in RTs.

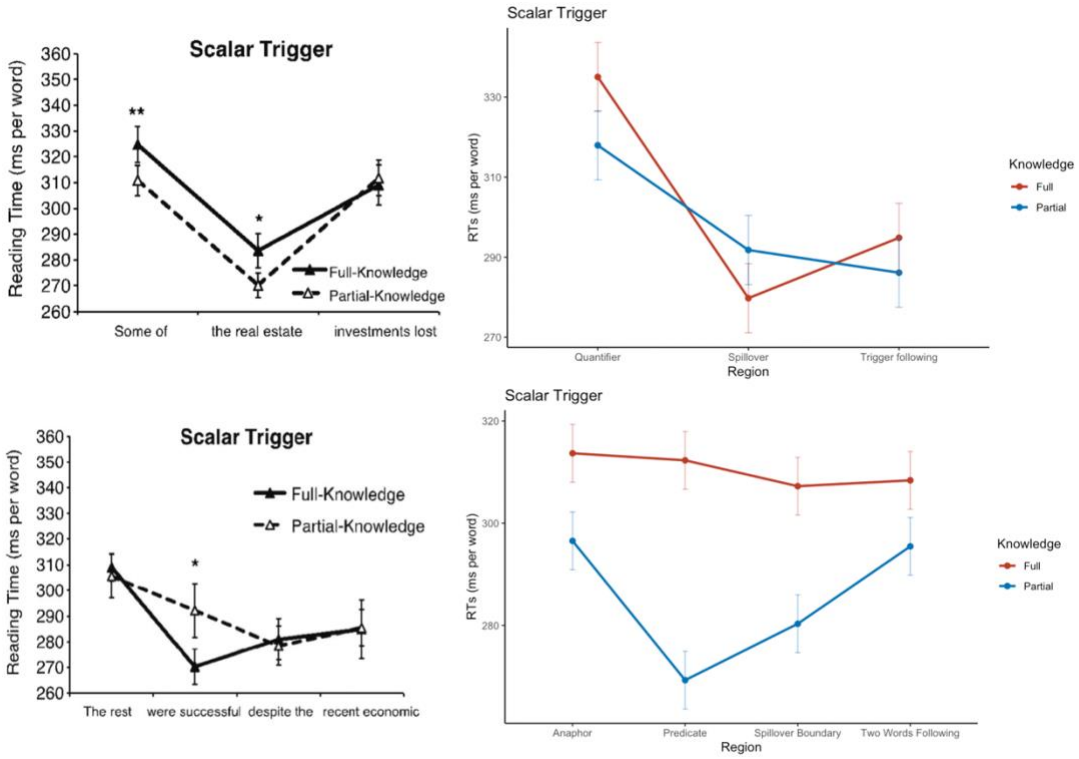
Linear Mixed-Effects Regression Models

As in the original study, I conducted linear mixed-effects regression models over the specified regions of interest in the trigger and complement sentences. The observed results are portrayed

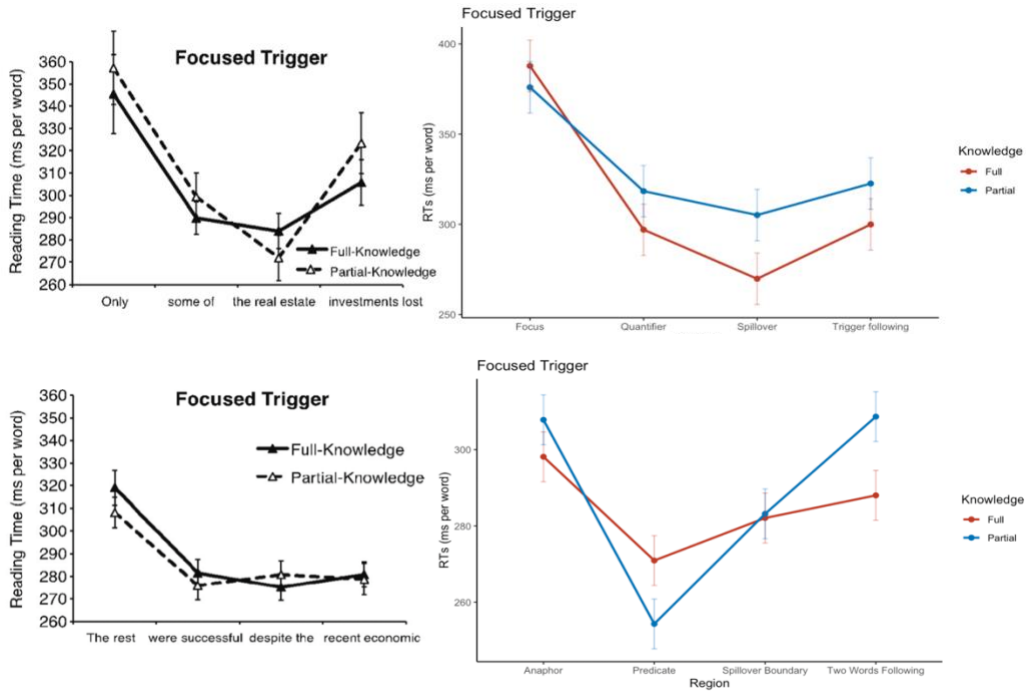
in Tables 2 – 5. To note, given the way R assigns baseline comparisons in models, positive coefficients indicate higher RTs within the partial-knowledge condition and negative coefficients indicate higher RTs within the full-knowledge condition. Similarly, positive coefficients regarding trigger type indicate higher RTs in the scalar trigger condition; negative coefficients indicate higher RTs in the focused trigger condition.

From Tables 2 and 3, we extracted multiple findings. First, we observed that the coefficient for the knowledge condition in the scalar quantifier region was positive, indicating increased RTs over the scalar quantifier in the partial-knowledge condition than in the full-knowledge condition. Even though the result is not significant (neither was the corresponding result in the original study), this contrast still contradicts our initial linking assumption that we would see increased reading times in the scalar quantifier region in the full-knowledge context. Beyond this discrepancy, we see that in the trigger sentences, my results replicated the general trend original findings, but with smaller effect sizes (as denoted by the significance levels). However, it is interesting to note that the one area where the effect contrasted from the original results is in the interaction of knowledge and trigger type in the scalar quantifier region.

Tables 4 and 5 illustrated the effects of knowledge and trigger type over the anaphor and predicate regions in the complement sentence. Our seconding linking assumption predicted lower reading times over the complement sentence in the full-knowledge context. The negative coefficients regarding the effect on RTs over the anaphor and predicate regions blatantly contradicted this assumption, even if the results were not significant (the corresponding findings in the original study were also insignificant). The original study stated that the results from the models for the complement sentences were insignificant (except for the interaction effect over the predicate region) without providing the actual coefficients, so it was difficult to compare the actual effects. However, all of my corresponding findings were also insignificant. The only discrepancy was in the interaction effect over the predicate region. Overall, my effects had the same level of significance as in the original study but contradicted the second linking hypothesis.



Figures 1 (top) and 2 (bottom). Depict average reading times over each region with 95% CI error bars for both knowledge contexts in the scalar trigger condition. Figure 1 is over the trigger sentence; Figure 2 is over the complement sentence.



Figures 3 (top) and 4 (bottom). Depict average reading times over each region with 95% CI error bars for both knowledge contexts in the focused trigger condition. Figure 3 is over the trigger sentence; Figure 4 is over the complement sentence.

Quantifier

	Coefficient	t-value	p-value
Knowledge	--	1.03	--
Knowledge	.008	.418	.676
Trigger	.035	4.98	.0001***
Trigger	.054	2.808	.005**
Interaction	.016	2.37	.0164*
Interaction	-.042	-1.081	.280

Spillover

	Coefficient	t-value	p-value
Knowledge	.015	2.49	< .05*
Knowledge	.069	3.126	.003**
Trigger	--	--	--
Trigger	.004	.168	0.868
Interaction	--	--	--
Interaction	-.048	-1.241	.218

Tables 1 (left) and 2 (right). Depict the effects and significance levels of knowledge and trigger type (and the interaction effect) on RTs in the trigger sentence. The top row in each effect depicts the original findings; the bottom row represents my findings. Dashes indicate the values the researchers did not give because they were insignificant. Table 1 shows the effect over the scalar quantifier; Table 2 shows the effect over the spillover region.

Anaphor

	Coefficient	t-value	p-value
Knowledge	--	--	--
Knowledge	-.049	-1.603	.109
Trigger	--	--	--
Trigger	-.041	-1.329	.184
Interaction	--	--	--
Interaction	-.085	-1.379	.168

Predicate

	Coefficient	t-value	p-value
Knowledge	--	--	--
Knowledge	-.058	-1.310	.205
Trigger	--	--	--
Trigger	-.047	-1.304	.196
Interaction	.018	2.09	< .05*
Interaction	-.109	-1.560	.122

Tables 3 (left) and 4 (right). Depict the effects and significance levels of knowledge and trigger type (and the interaction effect) on RTs in the complement sentence. The top row in each effect depicts the original findings; the bottom row represents my findings. Dashes indicate the values the researchers did not give because they were insignificant. Table 4 shows the effect over the anaphor; Table 4 shows the effect over the predicate.

Discussion

The primary motivation of the original 2012 study by Bergen & Grodner was to determine how, and when, perceived speaker knowledge level affects the generation and comprehension of scalar implicatures. The researchers concluded that speaker knowledge does affect the likelihood of generating a strong or weak implicature, using their findings regarding the 3 linking hypotheses to support this conclusion. Through their study, they found that full-knowledge increased RTs at the scalar quantifier (greater processing difficulty) in the trigger sentence and decreased RTs over the complement sentence. They also determined that speaker knowledge does not affect processing difficulty in the trigger and complement sentences in the focused trigger condition.

Through my replication, I reproduced mostly similar findings, albeit with lesser effects. My first series of results indicated that in the full knowledge context, RTs (and associated processing difficulty) increased over the scalar quantifier, but effect of knowledge context was insignificant. To this extent, these results replicate the original findings quite well. In contrast, my results from both the comparative and regression analyses for the second linking assumption showed higher complement sentence RTs in the full knowledge condition, even if insignificantly so (from the p-values in the regression analysis). These results are different form the original findings. I suspect this difference may be due to lack of power in the continuation sentences, as continuation sentences were split into complement and cancelation sentences. My final set of results followed the original study, with no significant difference in RTs over the trigger sentence quantifier and complement sentence anaphor. However, I only performed the comparative analysis and not the regression analysis for this hypothesis, indicating that my results are only a partial replication.

I have discussed the extent to which my findings match the results Bergen & Grodner found, as well as potential reasons for any discrepancies. Yet it's hard to fully compare the two sets of findings given the original researchers often excluded the actual effect coefficient. Thus, the comparison of my results with the original findings is limited to a certain extent. From my replication, despite discrepancies in my findings, I was able to reproduce the qualitative results Bergen & Grodner found. This overall reproducibility provides reassurance of their findings and reiterates the importance of considering speaker knowledge within conversational contexts.

References

1. Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(5), 1450–1460. doi: 10.1037/a0027850