How does reading experience mediate the extent to which entropy influences semantic plausibility?

Linette Anyiacha

Department of Psychology, University of Alberta

PSYCH 496: Individual Study

Dr. Christina Gagné & Dr. Thomas Spalding

April 24, 2025

Abstract

Language experience shapes how individuals process meaning, enabling greater precision in interpreting novel information. While reading experience plays a crucial role in this process, the mechanism by which it interacts with lexical properties remains unclear. This study examines how reading experience mediates the effect of entropy (semantic ambiguity) on plausibility judgments during compound word comprehension. Participants completed a Possible Relation Task (PRT) to assess their interpretation of compound words under different semantic priming conditions (related or unrelated), and an Author Recognition Task (ART) to measure reading experience. Results revealed that priming condition did not significantly affect response times or accuracy. Higher entropy was associated with longer response times and a lower likelihood of selecting the most frequent interpretation, while greater reading experience correlated with increased response times rather than faster comprehension. These findings suggest that reading experience fosters deeper semantic processing rather than expediting meaning resolution for high-entropy words. The ambiguity inherent in compound words remains a primary driver of comprehension difficulty, independent of immediate context. Rather than mitigating entropy's effects, reading experience may introduce an additional layer of complexity, encouraging deliberation and broadening interpretive possibilities. While it may "level the playing field" by ensuring consistent engagement across participants, it simultaneously presents another form of entropy that is rooted in cognitive effort rather than lexical frequency or transparency.

Introduction

Existing literature on semantic knowledge has often posited conceptual combination as the glue to our understanding of and utilization of complex stimuli structure that we encounter each day. Language experience taps into the ongoing sum of an individual's interaction with the world in various mediums. Narrowing our focus to only exposure to lexical information, we can look at how reading experience gives rise to a complex understanding of the world. Considering reading experience allows us to operationalize language experience within the constraints of an experiment, although many things can contribute to people's derivation of semantic plausibility. On a broader scope, semantic plausibility draws upon pre-existing knowledge and is then mediated by the process of conceptual combination. As defined by Gagne et al. (2005), conceptual combination is the process of putting together two or more concepts that do not, as yet, exist as independent representational structures.

Schmidtke et al. (2018) suggested that semantic knowledge about compounds becomes both more precise and flexible as a consequence of increased exposure to materials. The compound's comprehension is then facilitated by its constituents based on the individual's language experience. This facilitation by constituents was investigated in terms of semantic transparency, pertaining to the extent to which the head and modifier retained and contributed meaning in an overall compound word (Gagne et al., 2019). In considering all this, we must not neglect the role of context in the process of comprehension. In the construction of knowledge, people build upon their experiences, and context becomes a highly active subset of the person’s knowledge base (Gerrig & Murphy, 1992). Subsequently, any process that accesses information in the discourse context is necessarily knowledge-based. In a conversation for example, people would start discussing a particular topic. As the conversation gets more complex in its subject

and after many words have been exchanged, discoursers start "relating words and phrases to referents that have been previously established," which helps facilitate their comprehension and helps them understand the connected discourse (Gerrig & Murphy, 1992, p.207). This is what they described as relational continuity. Based on that reasoning, a priming condition among individuals with different reading and or language experiences would level the playing field and there'd be no observed differences in their understanding of meaning and word processing. In the third experiment in Gerrig and Murphy (1992), they gave subjects an opportunity to display the extent of their understanding of novel compounds within the context of self-contained stories. They found that context gave participants enough support to interpret the novel phrase, allowing easy integration of information with the rest of the story. In their investigation, they did not consider what aspect of semantic processing may be influenced by people's language experience, and rather focused on the result itself, prompting further investigation.

Gagne et al. (2005) suggest that past experience with various combined concepts gives rise to knowledge that can be used to determine the probability that a given concept will use a particular relation. The modifier concept's relational information usually plays a larger role in conceptual combination than does the head. That is, the first word in the compound played a larger role than the second. Thus, trying to understand conceptual combination in terms of independently represented relational information that is not connected to either of the constituents in a compound did not account for existing data at the time. Though not outright mentioning semantic plausibility, Gagne et al. (2005) discussion heavily imply the concept. These postulates paved way for subsequent extensive studies of conceptual combination in terms of semantic plausibility from several researchers, who operationalized it differently.

Christianson et al. (2010) operationalized it as the complex interaction between syntax and semantic routes of comprehension that would show differential activation on an ERP. While Gagne and Spalding (2013) did not directly mention semantic plausibility, they described how concepts are implicitly stored based on some underlying relation to one another, which when expressed linguistically, is omitted because it is assumed. Wang et al. (2018) operationalized semantic plausibility by injecting world knowledge into a computer model by creating physical events and considering various factors and characteristics of each. They described it as the recognition of events that are plausible, but possibly novel. Their model failed on highly specific attributes that humans would have understood, prompting the understanding that many attributes of meaning are not always explicitly coded, especially in terms of selectional preferences, which we often take for granted because they appear obvious. Mullaly et al. (2010) further investigated semantic plausibility in terms of the dominant sense of the particular stimuli – the most frequent interpretation of the word. This interpretation would be based on ratings in an established compound word database.

Common to the latter authors is the understanding individuals construct and refine their underlying, implicitly coded knowledge stores based on experience and in their interaction with the world. As a result, they engage in semantic processes often automatically; they just know it, do it, or say it. Guiding this automaticity is the influence of context. We do not pay attention to underlying linguistic concepts and take certain phrases for granted because they are always seemingly accessible to us. If we had to explain it, it'd be a lot harder than if we just did. This is one part of what the current project aims to pinpoint. We aim to understand that underlying point at which people start comprehending semantic information and just how much of their language experience is leveraged in the process.

In terms of how people arrive at meaning itself, Schmidtke et al. (2018) used the possible relations task where participants were presented with 16 possible relational meanings for a compound word but had to select the most plausible (i.e., the most correct) one. Their understanding highlighted the underlying factor, entropy, that would highlight the difficulties encountered during the processing of the compound. The interpretation that the participants eventually arrived at would be heavily dependent on their ability to construct a plausible concept (Gerrig & Murphy, 1992). Gagne et al. (2010) investigated meaning for phrases and did not necessarily use compound words, but rather components such as prevalence and goodness, which they thought would be better evaluated individually. They thought those components contributed to meaning predictability and ease of interpretation. Prevalence was a reading's accessibility, and goodness was a reading's plausibility. Prevalence was described "the likelihood that a reading will be provided by other language users," which echoes our current understanding of word frequency (Gagne et al., 2010, p.246). Goodness drew on various sources of knowledge in the world, and thus would affect the accessibility of meaning, which is the ease of processing. That closely aligns with current understanding of entropy, which is intertwined with frequency, as they direct meaning predictability and ultimately drive interpretation of information. As such, these factors are pertinent to not just the *what* aspect of interpretation, but the *how* aspect as well to understand with greater precision how those experiences contribute.

The previous finding by Gagne et al. (2010) laid the groundwork for Mullaly et al. (2010), who used the terms entropy and frequency explicitly in their paradigm, expanding on their previous definitions. They identified the frequency of a word's interpretation in operationalizing semantic plausibility, which added another dimension to understanding the comprehension process. Frequency, they thought, is tightly interwoven with entropy, where

entropy is the number of possible interpretations or meanings for a given word. High entropy indicates many possible plausible relational interpretations, while low entropy entails one dominant relational meaning. Implicitly, entropy highlights the ease or difficulty involved in processing information and can be measured in terms of reaction time. On the other hand, word frequency refers to how often a word appears in general language usage. Understanding this could potentially enhance our ability to precisely identify the degree to which language experience influences outcomes. Later investigation of the effects of language experience on the combinability of concepts in relational knowledge, Schmidtke et al. (2018) found that greater language experience resulted in more possible relational meanings, but greater precision about which of the meanings was the most semantically plausible. While this confirmed previous hypotheses about the influence of language experience on the quantity of possible relational interpretations, still, little was known about the degree of semantic plausibility of relational information.

Current Issue

Highlighting the degree of language experience influence entails focusing on the factors underlying the comprehension process that leads to greater precision in meaning. This is the aim of the current project, as we hope to pinpoint entropy as a key player in said process, leading to the question: how does reading experience mediate the extent to which entropy affects semantic plausibility? We hypothesize that individuals with more reading experience would be able to narrow down possible meanings and reach the dominant sense of compound word meaning much quicker, which would indicate that reading experience most directly affects comprehension earlier on in the process. Though Schmidtke et al. (2018) looked broadly at language experience, we intend to emphasize reading experience, as it can be better studied within an experimental

setting. Reading experience would be gauged in terms of exposure to written material which require its consumers to understand the semantic information. This information is never presented in a vacuum, as is information outside of the material's constraints. Based on extensive previous research in the literature, knowing that context is implicitly injected into many real-world conversations and instances of semantic digestion means prompts consideration in our proposed design. Context would then be considered the leveler of the playing field between individuals with high and low reading experience to gauge to what extent their experiences influence their ability to arrive at the dominant sense of various compounds.

METHODS

Study Design

This study employed a randomized counterbalancing design to investigate how reading experience mediates the extent to which entropy affects semantic plausibility. There were two possible experimental lists that participants could see. The lists were then randomized to ensure that the order of items was varied across participants. This randomization mitigates any potential order effects and ensures that each participant is exposed to a unique sequence of experimental items. To further reduce bias, the assignment of experimental items was randomized independently for each trial, preventing any systematic influence of item order or grouping on participants' responses.

Participants

The study included 216 undergraduate level students in PSYCH 104 and PSYCH 105. Participants were recruited using both convenience and self-selection sampling, as the

experiment was posted onto the University's psychological experiments page and participants decided to partake. 107 participants saw list one and 109 saw list two. All participants provided informed consent before participation.

Materials

The compound words used in the experiments came from existing compound word databases such as SUBTLEX and the Large Dictionary of English Compound words (LADEC), with existing frequency, entropy, and semantic transparency ratings.

The experimental items listed consists of 64 compound words selected from the databases within the following criteria: we considered frequency as well, and the values ranged from 0.3 to 2.5, as to include words that were neither too frequent nor too rare. Entropy was selected more intentionally within bounds of 1.5 and 3.5, and the frequency just came along with it. Frequency and entropy are inversely correlated anyway, and their values for each compound reflected this relationship. The greater entropy bounds satisfy the experimental goals of investigating how entropy (presumably high entropy) affects semantic plausibility. However, the general rationale was that the selected words needed to be those that everyone would be familiar with to an extent. The chosen items did not have more than one semantic sense (i.e., they did not have both a verb and noun meaning). See the appendix for the experimental item list.

The experimental components were built into and run on Gorilla software (Gorilla Experiment Builder, n.d.). STATA 18 and JASP 0.19.10 were used for data analysis (StataCorp, 2023; JASP Team, 2024).

It was determined earlier on that 3 different tasks on the experiment would be too tedious for participants, and therefore in the analysis, I used existing semantic transparency ratings to see

whether the modifier component meaningfully affects reaction time and ultimately what relation was chosen. The semantic transparency ratings were used during the analysis to supplement understanding. The final experimental design consisted of 2 tasks: the Possible Relation Task (PRT) and the Author Recognition Task (ART).

Procedure

Part 1: Possible Relation Task

Before the task began, participants were presented with an introduction outlining general information of the experiment, and then the ethical considerations and consent. If consent was given, participants proceeded to the task. Because the experiment flow included a built-in randomizer, participants were assigned to either list one or list two, where the items were presented in different orders in a counterbalanced manner. This ensured that each item was paired with only one of the primes, either related or unrelated. This was the first task designed to prevent participants from guessing the experiment's hypothesis.

Before the participants saw the first compound, they went through one practice trial that demonstrated the flow of the task. The flow was as follows:

Ready? > Fixation > Prime > Question "what is the MOST LIKELY meaning?"> Compound with the possible relations

"Ready" appeared on the screen and remained until the participant clicked on or pressed a button. This was followed by a fixation point, which was a black cross hair, on the top of the screen, which was the position in which the prime would appear. The prime then appeared for 100 ms and was followed by the compound words and 16 possible relations. After the participant completed the single test trial, the 64 experimental items appeared in the same manner.

From this task, entropy will be calculated based on the participants' reaction times in terms of how long it took them to select a relational meaning for each of the compounds, and by extension, the difficulty they experienced in the process. As well, the dominant relational sense of each compound will be considered in terms of the maximum proportion of people that choose any one relational meaning. The analysis of the dependent variable, response time, was analyzed based on average reaction times for each of the compounds. The analysis consists of linear regressions investigating whether semantic transparency rating of the modifier constituent, entropy, and the priming condition contribute meaningfully to the resulting reaction times.

Part 2: Author Recognition Task

This task begins right after the Possible Relation Task (PRT), and comprises a list of names of authors, both real and made up. This task was used to gauge reading experience through exposure to print material.

The participants were instructed as follows: Below is a list of names. Some of them are authors of books, and some of them are not. Please select the ones that you know for sure are authors. There is a penalty for guessing, so you should check only those names about which you are absolutely certain. Thank you. Scroll down to see all the options and once you have made all your selections, please click "Next". Of this list, 65 of the names are real authors, and therefore 65 is the highest participants can score. For this task, there was a penalty for incorrectly selecting a name that wasn't an author name. One point was awarded for each name selected correctly, and one point was lost for each name that was incorrectly selected. The analysis of this task will be considered individually across the 216 participants.

Results

This study investigates how reading experience mediates entropy's influence on semantic plausibility, through the PRT and the ART. The former places emphasis on semantic transparency and entropy as they affect comprehension under varying contextual conditions, while the latter places emphasis on language exposure. The current preliminary analysis focuses on response times across the different priming conditions. The analysis includes averaged response times (RTs) from 216 participants.

Participants RTs were analyzed across the related and unrelated priming conditions. The mean RT for the related condition was *M* = 12035 ms (*SD* = 5235, range = 4948-38818) and a median of 10400 ms. The mean RT for the unrelated condition was M = 11587 ms (*SD* = 3978, range = 5572-25343) and a median of 11203 ms. The RTs in the related and unrelated priming conditions appear to be moderately positively skewed, with skewness values of 2.351 (*SE* = .299) and 1.105 (*SE* = .299), respectively.

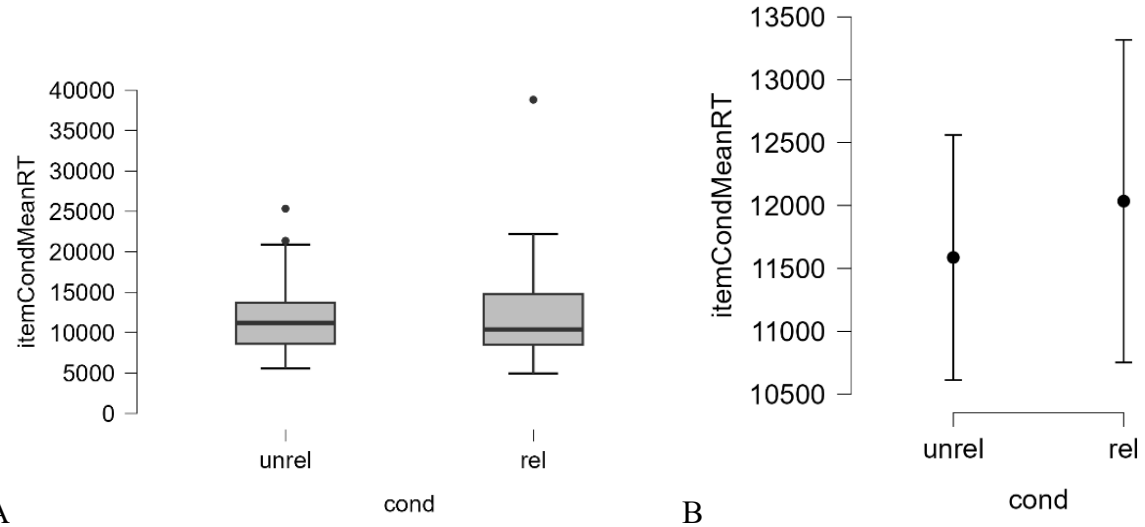

Figure 1: Graph A is a boxplot of response times (RTs) by condition, showing two outliers in the unrelated (unrel) condition and one outlier in the related (rel) condition. Graph B is an interval

plot of mean response times (RTs) by condition, showing longer response times and wider variability in the related (rel) condition compared to the unrelated (unrel) condition.

Although there are three outliers overall, as seen in Figure 1A, it is important to note that the response times (RTs) are averages across participants. Removing any single data point could potentially alter the overall pattern, as each outlier reflects the group average for a specific set of participants' responses. The seemingly longer response times in the related condition, in Figure 1B, suggest that exposure to a related prime may prompt participants to spend more time considering how the prime connects to the compound's relational meaning. This could indicate a cognitive shift from the straightforward semantic association elicited by the prime to the more abstract task of identifying a relational interpretation. The added processing time might reflect the effort required to reconcile the prime's direct semantic meaning with the relational structure of the compound word. However, these differences were superficial, as a t-test revealed no significant differences. A Mann-Whitney test confirmed that the mean RT in the related prime condition did not significantly differ from that of the unrelated prime condition ($U = 2019$, $p = .892$). The effect size was small ($rB = -0.014$), suggesting a negligible difference between the priming groups.

A multiple linear regression model was used to examine the effects of priming condition and semantic transparency ratings on RT. The fitted regression model was: Mean RT = 11068.675 + 57.643*(semantictransparencyC1) + 454.479*(priming condition) – 8629.770*(maxProportion). The overall regression was statistically significant ($R^2 = .14$, $F(3, 124) = 6.610$, $p<.001$). Semantic transparency ($\beta = 57.643$, $p=.016$) and maxProportion ($\beta = -8629.770$, $p<.001$) significantly predicted mean RT, but not priming condition ($\beta = 454.479$, $p=.556$). Based on the model, as semantic transparency of the first constituent increases, so does RT. It does not seem to

have an effect on its own. As maxProportion increases, RT decreases, that is, the more people agree on one relation, the more the RT decreases. Priming condition was not a good predictor of RT. So, whether the participants were in the related or unrelated priming condition did not affect their choices or their RT. This indicates that relational choices were stable, meaning that the underlying process for choosing and interpreting the relations between compounds remained consistent, regardless of the type of prime they were exposed to. In other words, participants' relational interpretations were not swayed by the priming condition, indicating that their decision-making process was steady and reliable across different contexts. Semantic transparency does not have an effect on its own, seen when included alone in a regression model.

Another multiple regression model was used to examine the effects of entropy and priming on RT. The fitted regression model was: Mean RT = 4770.342 + 419.336*(priming condition) + 2370.066*(entropy). The overall regression was statistically significant (R^2 = .091, F(2, 125) = 6.278, p=.003). It was found that entropy (β = 2370.066, p<.001) significantly predicted mean RT, but once again, not priming condition (β = 419.336, p=.595). When I set up the model with entropy and maxProportion or diversity and maxProportion at the same time, they both appeared to not be good predictors of mean RT. Supposing that this was due to their close correlation, I investigated further through correlation coefficients.

Pearson correlation coefficients were computed to assess several linear relationships within the study. There was a negative correlation between maxProportion and entropy, r(214) = -.967, p<.001. There was also a negative correlation between maxProportion and diversity, r(214) = -.529, p<.001, where diversity is the number of relations selected for any given compound. Across the board with the data, diversity is high for all the compounds, indicating less consistency with relational interpretation, and this is what is expected when entropy is at play.

A Pearson correlation coefficient was calculated to examine the linear relationship between word frequency and mean response time (RT). The relationship was not significant, $r(214) = .156$, $p = .078$. Because word frequency is correlated with entropy and the frequency values ranged widely (0.3 to 1.5)—reflecting typical variation in everyday language use—it is possible that entropy already accounted for much of the variance, leaving word frequency with no significant independent effect.

Finally, there was a significant positive correlation between entropy and mean response time (RT), $r(214) = .299$, $p<.001$, indicating that higher entropy is associated with longer times for participants to choose a relational interpretation. However, the modest correlation suggests that entropy only accounts for 8.9% of the variance in RT. This finding underscores that many other factors likely contribute to the process of interpreting the meaning of compound words.

A linear mixed-effects model was fit to examine how participants' reading experience (as indicated by their ART scores) mediates the effect of entropy on semantic plausibility during a priming task. The model included RT as the dependent variable, priming condition (related/unrelated), ART score and entropy as fixed effects. Random intercepts were included for subjects and items to account for individual differences in response times and item-specific variability. There appeared to be no interaction between the priming condition and the entropy score, so the simpler model was used.

The model showed that entropy significantly increased RT ($\beta = 2274.925$, $SE = 806.239$, $t = 2.822$, $p = 0.006$), indicating that higher entropy led to longer response times. Priming condition did not significantly affect RT ($\beta = -418.305$, $SE = 296.549$, $t = -1.411$, $p = 0.158$). Curiously, ART score significantly increased RT ($\beta = 234.295$, $SE = 79.641$, $t = 2.942$, $p = 0.004$), suggesting that participants with higher ART scores responded slower.

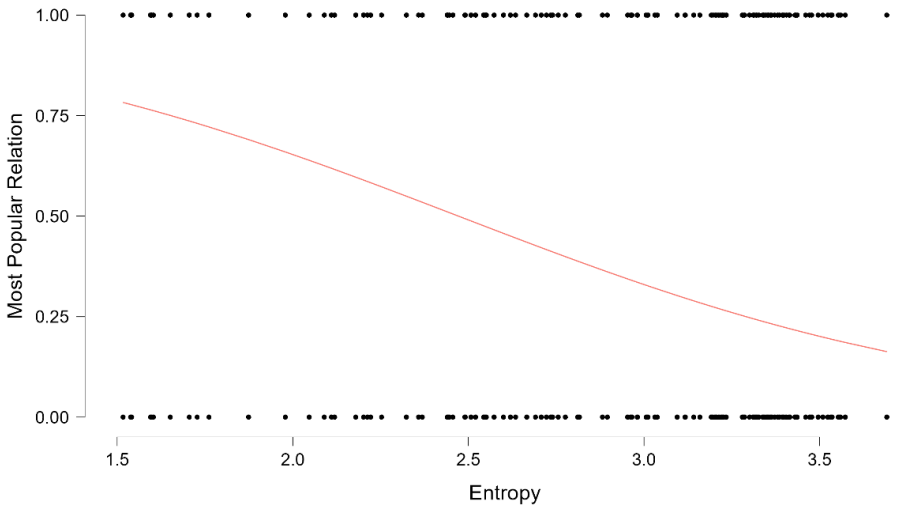

Figure 2: A graph depicting the negative relationship between entropy and accuracy (in this case selecting the most popular semantic relational interpretation).

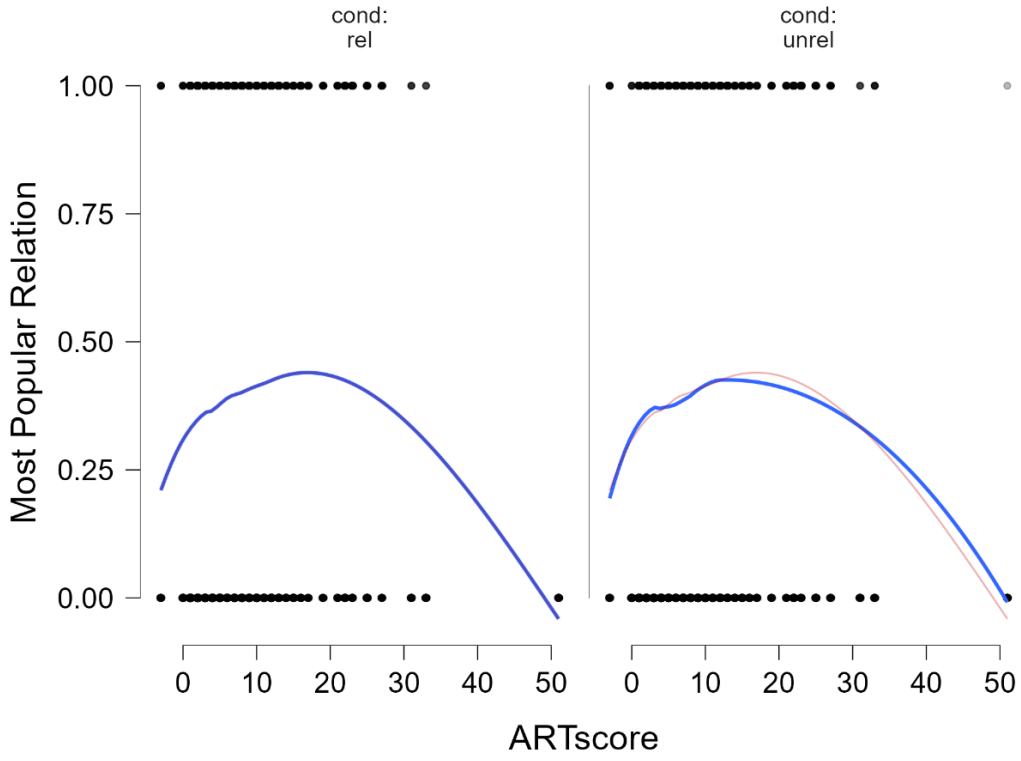

Figure 3: Graph depicting priming condition and ART score as they predict likelihood of selecting the most popular semantic relation.

A generalized linear mixed model was fit to examine how entropy, priming condition and ART score predict the accuracy of participants in selecting the most popular (and thus, the most plausible) semantic relation. The fixed effects of entropy, priming condition, and ART score were tested, with subject and stimuli included as random effects to account for individual and item-level variability. The binary outcome variable was the selection of the most popular relation (coded as 1 for correct and 0 for incorrect) and thus the model was run as a binomial family with a logit link.

The ANOVA summary showed that entropy ($\chi^2(1) = 199.939$, $p < .001$) significantly predicted accuracy, as shown in figure 2. ART score ($\chi^2(1) = 0.121$, $p = 0.728$) and priming condition ($\chi^2(1) = 0.030$, $p = 0.862$) were not significant predictors. Figure 3 shows clearly how higher ART scorers were less likely to select the dominant relational interpretation. The inverse u-shape suggests that the mid-level scorers are more likely to select the popular relation, while the lower scorers are less likely, and the higher scorers are the least likely to do so. The lower scorers may have relied more on guessing or may have defaulted to simpler associations, while the higher scorers could have been more consciously analytic of each compound and considered more alternative, nuanced, and less common relational interpretations. In the middle, the mid-range scorers likely had enough reading experience to recognize dominant relations without overanalyzing them. Thus, they may have followed conventional interpretations that aligned with their knowledge.

Discussion

Across all the different types of analysis, from the linear regression to the linear mixed effects models, priming condition (whether a semantically related or unrelated word preceded the compound) did not significantly impact response times or the likelihood of selecting the most popular semantic relation. This disagrees with the findings of Gerrig and Murphy (1992), where a contextual cue provided enough support for people to interpret the stories they had been presented with. This context effect has been shown in other studies, so perhaps the null effect of priming in my investigation may be due to something particular to my implementation of it, and may even suggest that something beyond immediate environment plays a larger role. Although Gagne et al. (2019) demonstrated that the modifier constituent retains and contributes to a compound word's overall meaning, participants in this study may have processed each compound holistically, rather than decomposing them into their modifier and head components. In this case, priming at the constituent level may not have had a strong impact because participants were not engaging with the individual parts of the word separately. Another limitation in implementing may have been the duration of the prime. The priming window of 100 ms may have been too short or weak to be seen or even processed by the participants, and therefore the participants could have just been interpreting the compound word based on their prior knowledge anyway. Unlike single-word priming effects commonly found in lexical decision tasks, compound word comprehension might rely more on lexical properties such as entropy, frequency and transparency than on brief semantic priming.

The Just & Carpenter (1992) capacity theory provides a useful framework for understanding the relationship between activation, semantic priming, and comprehension. In this model, activation is a finite resource shared between storage and processing in working memory. The nature of the primes in my study may not have been sufficient to fully activate relevant semantic

representations, limiting their ability to facilitate comprehension. This aligns with capacity theory's premise that when activation is not effectively allocated to task-relevant elements, comprehension may be constrained, especially in tasks requiring semantic plausibility judgments.

The finding that participants with more reading experience had longer reaction times (RTs) offers additional insight into the role of individual differences. Rather than prioritizing speed, experienced readers appear to engage in more effortful processing, integrating multiple interpretations before making a decision. Their longer RTs likely reflect a deliberate allocation of activation to support the processing of multiple relational interpretations rather than a simple retrieval of the most accessible meaning. This supports the idea that individual differences in working memory capacity shape not only the amount of activation available but also the strategies used to allocate these resources.

The failure of priming to enhance comprehension for semantically plausible interpretations highlights another dimension of capacity theory: the trade-off between processing and storage. In the current study, activation from priming may have been misaligned with the specific demands of constructing plausible semantic relationships, reducing its efficacy. Overall, this suggests that while semantic priming can preactivate elements in working memory, its effectiveness depends on both the strength of the primes and how well the activation aligns with the task's cognitive demands.

The effect of language experience (gauged by reading experience using the ART), while significant, turned out to be in the opposite direction from what I hypothesized. Based on the linear mixed effect model, there is a positive relationship between reading experience and reaction time, such that as ART score (i.e., reading experience) increases, so does reaction time.

But what underlying process could be taking place? Those with greater reading experience may be undergoing deeper semantic processing, longer deliberation due to greater cognitive flexibility, and even familiarity with word structures. Individuals with higher ART scores may engage in deeper semantic processing thanks to their familiarity with language and word meanings. This would make them take longer to decide on the most semantically plausible interpretation since they could be considering multiple meanings before making a decision. They could be weighing multiple factors such as the frequency and nuances of meanings or associations between the components of the compound word before settling on the most plausible interpretation.

Entropy's positive relationship with response time was more in line with what has been shown time and again in the literature. The positive relationship suggests that as the ambiguity or uncertainty of the compound word increases, individuals require more time to process and determine the most semantically plausible meaning. Higher entropy compounds, which may have multiple potential interpretations, increase the cognitive load involved in resolving this uncertainty. As a result, individuals take longer to process these words and make their decisions, leading to increased RT. This highlights the role of semantic complexity, and the additional time needed to evaluate multiple possible meanings when presented with higher-entropy compounds.

Now, to answer the question of how reading experience mediates the effect of entropy on semantic plausibility: reading experience does not mediate entropy's effects on plausibility judgments in the expected way (i.e., it does not reduce entropy's impact). The fact that ART score increased RT suggests that reading experience does not reduce the cognitive effort required to process high-entropy compounds. Reading experience might "level the playing field" across participants, ensuring they engage with the task similarly, but it also introduces a different kind

of entropy related to deliberation rather than lexical frequency or transparency. During increased deliberation, more experienced readers might be more aware of multiple possible semantic relations, leading them to think more deeply about the compounds rather than defaulting to quick, intuitive judgments. In other words, ART score doesn't just reflect reading experience, but also signal a tendency toward greater linguistic analysis, which in turn slows down decision-making.

These findings provide a deeper understanding of the mechanisms underlying compound word comprehension, potentially pointing to a two-stage processing model. First, a compound word's inherent entropy may increase cognitive load, making meaning extraction more challenging. Then, reading experience may modulate the depth of processing by introducing its own form of entropy, leading to longer deliberation rather than a straightforward facilitation effect. Thus, experienced readers engage in more effortful but flexible approach to meaning resolution rather than automatically overriding ambiguity with prior knowledge.

I initially hypothesized that context (priming) might mitigate the effects of entropy, making less predictable compounds easier to process. However, priming did not "level the playing field" by reducing processing differences across reading experience levels; rather, it ensured that participants engage similarly with the task. Its lack of significant effect suggests that entropy exerts the strongest influence on semantic plausibility judgments, independent of short-term contextual facilitation. This was shown by the GLMM, where entropy was the only significant predictor of selecting the most popular interpretation, indicating that the inherent uncertainty of the compound words primarily drove plausibility judgments. Thus, entropy remains the dominant factor in decision making. More experienced readers may not necessarily be better at

resolving entropy, but may approach compound word comprehension more analytically, which paradoxically increases response time instead of decreasing it.

Finally, for simplicity, the current study did not deeply examine interactions between variables. When included, these interactions led to model overfitting and were all statistically insignificant. This may be due to factors such as the priming condition and sample size, which could be addressed in future research by increasing both. A larger sample and stronger priming manipulations might allow the model to detect meaningful interactions more reliably. Additionally, ART scores varied considerably and reflected individual differences that could have been accounted for as random effects. However, including them in the linear mixed models introduced excessive complexity and overfitting, necessitating their removal. To better understand why higher ART scorers had longer reaction times, future studies could incorporate a post-task questionnaire to capture participants' thought processes; whether they were considering multiple interpretations or were simply unsure of the most correct answer. This would help to distinguish between different levels of those with greater reading experience and identify the threshold at which reading experience becomes more cognitively costly.

Conclusion

These findings contribute to our understanding of how reading experience influences lexical ambiguity resolution, suggesting that experienced readers engage in a more effortful, deliberative approach rather than quickly resolving meaning based on prior knowledge. This supports models of language processing that emphasize depth of semantic engagement rather than mere efficiency. This has significant implications for educational assessment and reading instruction.

Traditional standardized tests often assume that stronger readers will perform better under time constraints, yet my findings suggest that highly experienced readers may take longer precisely because they are considering multiple possible meanings before arriving at a judgment. This raises questions about whether speed should be the primary measure of reading proficiency in standardized tests. Instead, assessments could be redesigned to evaluate the depth of comprehension, ambiguity resolution skills, and cognitive flexibility in interpretation rather than just accuracy under time pressure.

Additionally, reading instruction may need to emphasize not just vocabulary expansion and fluency but also strategies for efficiently managing linguistic ambiguity. For example, students could be trained to recognize when deliberation is helpful versus when it might slow them down unnecessarily. This could be particularly useful in subjects like literature, where interpreting multiple meanings is valuable, and in disciplines like law or philosophy, where precise meaning extraction is crucial. These findings also suggest a need to rethink the way reading ability is measured and taught, ensuring that assessments reflect the complexities of real-world language comprehension rather than an oversimplified view of reading as a linear, speed-based skill.

Ultimately, the findings lend to the idea that knowing more is costly: the more you know, the less you may really know. Again, this paradox suggests that more reading experience does not necessarily streamline comprehension, but instead encourages more effortful engagement with meaning. For now, it remains to be seen at what point knowing more shifts from being beneficial to burdensome. Future studies could further explore how different levels of reading experience influence cognitive load and develop strategies to help experienced readers leverage their knowledge more efficiently according to their specific needs, navigating this cognitive trade-off in a way that best supports comprehension.

References

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, 40, 278-289.

Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 538–544. https://doi.org/10.1037/a0018027

Gagné, C. L., Marchak, K. A., & Spalding, T. L. (2010). Meaning predictability and compound interpretation: A psycholinguistic investigation. *Word Structure*, *3*(2), 234–251. https://doi.org/10.3366/word.2010.0006

Gagné, C. L., Spalding, T. L., & Ji, H. (2005). Re-examining evidence for the use of independent relational representations during conceptual combination☆. *Journal of Memory and Language*, *53*(3), 445–455. https://doi.org/10.1016/j.jml.2005.03.006

Gagné, C. L., Spalding, T. L., & Schmidtke, D. (2019). LADEC: The Large Database of English Compounds. *Behavior Research Methods*, *51*(5), 2152–2179. https://doi.org/10.3758/s13428-019-01282-6

Gerrig, R. J., & Murphy, G. L. (1992). Contextual influences on the comprehension of complex concepts. *Language and Cognitive Processes*, *7*(3–4), 205–230. https://doi.org/10.1080/01690969208409385

Gorilla Experiment Builder. (n.d.). *Gorilla* [Computer software]. https://gorilla.sc

JASP Team. (2024). *JASP (Version 0.19.10) [Computer software]*. https://jasp-stats.org/

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149. https://doi.org/10.1037//0033-295x.99.1.122

Mullaly, A. C., Gagné, C. L., Spalding, T. L., & Marchak, K. A. (2010). Examining ambiguous adjectives in adjective-noun phrases: Evidence for representation as a shared core-meaning with sense specialization. *The Mental Lexicon*, *5*(1), 87–114. https://doi.org/10.1075/ml.5.1.04mul

Schmidtke, D., Gagné, C. L., Kuperman, V., & Spalding, T. L. (2018). Language experience shapes relational knowledge of compound words. *Psychonomic Bulletin & Review*, *25*(4), 1468–1487. https://doi.org/10.3758/s13423-018-1478-x

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 24, 402-433.

StataCorp. (2023). *Stata statistical software: Release 18*. StataCorp LLC.

Wang, S., Durrett, G., & Erk, K. (2018). *Modeling Semantic Plausibility by Injecting World Knowledge* (arXiv:1804.00619). arXiv. http://arxiv.org/abs/1804.00619