Recovering Meaning from "Meaningless" Texts: Semantic Ambiguity Reduction as Recovery of Distributional Structure

Kira Breeden*
kbreeden@wisc.edu
Psychology, University of Wisconsin-Madison

Gary Lupyan
lupyan@wisc.edu
Psychology, University of Wisconsin-Madison

Abstract

What does "blalfed" mean in "The fuite twouch blalfed on the splelve snaitch?" Although it may seem impossible to know, people *can* infer word meanings from such "nonsensical" contexts. The correct word–"sat"–was guessed correctly by 25% of our participants. We investigate people's ability to infer word meanings in such Jabberwocky texts. Participants guessed the meaning of a target word and selectively unmasked Jabberwocky words, allowing us to study the links between ambiguity, sampling choices, and accuracy of identifying the meaning of a target "nonsense" word. Ambiguity predicted how participants sampled the text and how successfully they recover meaning, with bidirectional entropy being more predictive than entropy that only considers information preceding the target word. Accuracy increased with early information gain. These results suggest that people combine structural expectations with strategic sampling, and that semantic information is often distributed across a passage rather than localized to a single cue.

Keywords: language understanding; morphosyntax; semantic ambiguity; information gain

Introduction

People frequently encounter new words whose meanings must be inferred from context. This context can take multiple forms including real-world referents such as objects or actions in plain view, social cues such as where the speaker is looking or pointing, as well as the surrounding linguistic context in which the word appears (Yu & Smith, 2007; Frank & Goodman, 2014; Gillette, Gleitman, Gleitman, & Lederer, 1999). Although direct referent–word pairings such as "look at the dog!" characterize some child-directed speech (Gentner, 1982), much of vocabulary growth in both childhood and adulthood relies on inferring meaning from language itself (Nagy, Herman, & Anderson, 1985; Boleda, 2020).

Studies investigating people's ability to use linguistic context as a cue to word meaning tend to embed novel words in highly informative contexts. For example, participants might be asked to infer the meaning of "draimant" given the context "Being unable to use it, the cooks always dumped a bunch of draimant in the alley behind the restaurant" (Borman & Lupyan, 2024), see also (Nagy et al., 1985). However, it has been recently discovered that it is, in principle, possible to recover meaning from linguistic contexts that are much more degraded. For example, large language models (LLMs) can often recover the full meaning of "Jabberwocky" sentences. Presented with the seemingly nonsensical input such as "At the ghybe of the swuint, we are haiveed to Wourge Phreargwurr, who sproles into a ghitch flount with his crurp", LLMs can recover meaningful English text, e.g., "At the start of the story, we meet a man, Chow, who moves into an apartment building with his wife" – a gloss that is very close to the original text (Lupyan & Yang, 2026; Lupyan & Arcas, 2026).

We have known for some time that people can infer various aspects of meaning from linguistic context, beginning with whether a word denotes an event or substance (and whether that substance is countable) (Brown, 1957) – a phenomenon later termed "syntactic bootstrapping" (Landau & Gleitman, 1985; Gleitman, 1990). Later work on construction grammars expanded the types of meaning that can be inferred from linguistic context to abstract properties like causation and transfer (Goldberg, 1995; Croft, 2001; Diessel, 2019). However, the recent evidence that LLMs are capable of recovering far more specific meanings from far more degraded linguistic contexts (Lupyan & Yang, 2026; Lupyan & Arcas, 2026) raise the question of whether people may be more capable at using linguistic context to infer meaning than expected by these theoretical accounts (De Luca & Lupyan, 2019).

Here, we examine people's ability to infer the meaning of words from degraded contexts, e.g., figuring out what "blalfed" means in "The fuite twouch blalfed on the splelve snaitch." In this scenario, one explanation for how people infer meaning is through syntactic bootstrapping (Gleitman, 1990; Landau & Gleitman, 1985). People exploit morphosyntactic cues—such as argument structure, tense marking, and pronouns—to infer semantic properties of novel words. Even very young children interpret novel verbs as causative when they appear in transitive frames and non-causative when they appear in intransitive frames (Naigles & Swensen, 2007). Applied to the sentence "The fuite twouch blalfed on the splelve snaitch," syntactic bootstrapping explains how readers identify which word is the agent, which denotes an action, and which is the patient.

We propose that in such impoverished contexts, meaning inference cannot be explained by syntactic structure or even linguistic constructions alone. Instead, it depends on broader forms of pattern matching over the entire distributional structure of language, where language users bring expectations derived from accumulated experience with how words, constructions, and events tend to co-occur. These expectations further constrain the space of plausible meanings.

We know that people can infer word meanings by exploiting patterns of co-occurrence (Aslin, 2017; Yu & Smith, 2007; Saffran, Aslin, & Newport, 1996) and contextual similarity (Goldberg, 2003; Nielsen & Christiansen, 2025). From this perspective, inferring meanings is a multi-cue constraint-satisfaction problem (MacDonald, 1994; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995): syntactic bootstrapping helps identify roles and relations, but distributional knowledge helps determine *which semantic hypotheses* are most compatible with prior linguistic experience. What is less clear is how people navigate this space when semantic cues are sparse and distributed across context rather than localized to specific words. In the extreme case, where all content words in a passage are replaced with nonce forms, semantic ambiguity is maximized by design and information relevant to a target word's meaning must be recovered from distributional structure. Although syntactic structure constrains interpretation to some extent, reducing ambiguity likely depends on multiple interacting factors, including the position of the target word, the structure of the surrounding context, and how much context becomes available over time.

To address this question, we presented participants with passages of text in which all content words were replaced with nonce words. Participants were tasked with guessing the meaning of one of the nonce words, and could incrementally reveal the words they thought would help them determine the target meaning. This allowed us to investigate how underlying semantic ambiguity interacted with distributional properties of the passage, including the position of the target word, the ambiguity in the context preceding and following the target, and the proportion of words participants chose to reveal in the passage. By examining both sampling behavior and guess accuracy, we evaluate how linguistic structure informs ambiguity and the recovery of meaning.

Methods

Participants were asked to infer the meaning of a nonce target word embedded in short passages as accurately as possible while revealing the fewest words. All content words in each passage were initially masked (replaced with nonce forms). Participants could reveal individual words (which cost them points), opting at any point to guess the target word. Guesses were restricted to single words. The task was designed to capture how participants sample contextual information under uncertainty and how this sampling behavior reduces semantic ambiguity.

Participants

We recruited two separate participant groups : a sampling group (n=160), who could chose which words to reveal, and a baseline group (n=80), who were given fully masked and fully unmasked passages, to compute baseline performance under minimal and maximal information conditions. All participants were recruited from Wisconsin-Madison's psychology participant pool in exchange for course credit. Participants were required to spend at least 4 seconds on each word-guessing attempt.

Materials

We constructed 80 2-3 sentence long passages using the base version of *gemma-2-27b*. There were 20 unique target words. For each word, we created four passage variants designed to manipulate how predictable the word was from the context (hence, 80 total passages). Target words spanned multiple parts of speech (17 nouns, 2 verbs, and 1 adjective) and appeared at variable positions within passages. Part-of-speech effects were unbalanced due to constraints imposed by the ambiguity manipulation.

We manipulated contextual ambiguity by creating passages differing in entropy over the model-predicted probability distribution of the target word, conditioned on the cumulative surprisal of the preceding context. For each target word, we first computed entropy values for a larger pool of candidate variants and then selected four variants that (a) shared the same most probable target word and (b) spanned a range of entropy values while preserving the intended meaning. For example, four variants describing an artist painting a landscape all favored the target word "hues" as the modal prediction, but differed in left-to-right entropy as a function of the context. Across passages, entropy values varied substantially ($M = 2.6$, $Mdn = 2.62$, $SD = 1.29$), though some target words were inherently more constraining than others.

In experimental trials, all content words except the target word were replaced with nonce forms from the ARC nonword database (Rastle, Harrington, & Coltheart, 2002). Each nonce word was "revealable"; upon clicking the target word, it was replaced with its original English word. The target word itself was in bold and could not be revealed.

Procedure

Sampling Group. Participants completed 20 trials. On each trial, participants viewed a passage with all content words masked and were instructed to infer the meaning of the target word displayed in bold text. Participants could reveal as many words as they wished before responding, but were instructed to submit a guess only when they could narrow their response to a single English word. To discourage participants from indiscriminately revealing all the words prior to their guess, participants earned points inversely proportional to the number of words they revealed.

Each participant saw one variant per passage (i.e., one entropy level per target word). Entropy varied within participants, but each target word was seen only once by each participant. Variants were counterbalanced such that exposure to higher-entropy passages for some target words was offset by lower-entropy passages for others. Passage order was randomized across participants.

Baseline Group. Participants in the baseline condition completed the same inference task *without* the option to reveal words and *without* point incentives. On each trial, par-

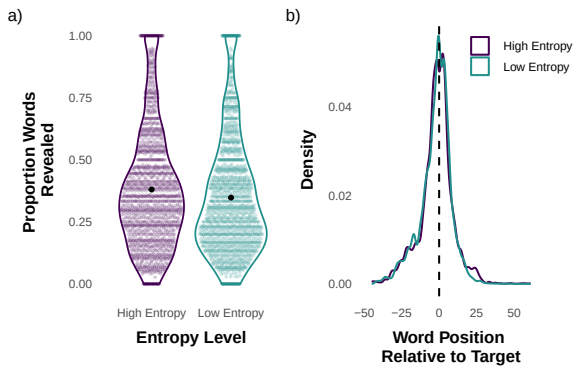

Figure 1: (a.) The relationship between entropy and proportion of words revealed. Participants revealed slightly more words for higher entropy passages. (b.) The relationship between entropy and how close revealed words are to the target word. People revealed words in roughly the same locations relative to the target in both higher and lower entropy passages. Entropy is binned (median split) for display purposes only.

ticipants read the passage, took as much time as needed, and then provided a single-word guess for the target word. Each participant saw 10 passages with all content words masked and 10 passages with all content words unmasked, except for the target word. Assignment to masking condition was counterbalanced across passages, and trial order was randomized. As in the sampling condition, participants only ever saw one passage per target word and entropy was counterbalanced as a between-subject manipulation.

Both Groups In both groups, after guessing the meaning of the target word, participants provided a confidence score on a 5-point Likert scale, indicating how confident they were that their guess was correct. The trial ended with a screen revealing the correct target word.

Results

The results are organized around three main questions: (1) How much and what kinds of information do participants solicit before guessing the target and how accurate are their guesses? Are some participants better than others? (2) How does ambiguity and information-seeking behavior jointly predict accuracy? (3) How are guessing accuracy and choices of which words to reveal influenced by different measures of semantic uncertainty? As these predictors lie on different scales, all variables in the following statistical models have been normalized.

How much context do people need to make a guess?

We first asked how much contextual information participants seek before guessing, and whether this behavior is related to our measures of ambiguity. To the extent that revealing words reduces ambiguity, the amount of revealed context and the locations of revealed words should reflect properties of the underlying ambiguity structure.

On average, participants revealed approximately six words before guessing the meaning of the target word ($M = 6.1$ words, $Mdn = 5$ words, $SD = 4.12$). Participants preferred to reveal words within a roughly seven word window of the target ($M_{distance} = 7.57$ words, $Mdn = 5$ words, $SD = 7.58$; see Figure 1). This pattern suggests that participants preferentially seek local contextual information rather than sampling uniformly across the passage.

But what influenced the amount of context people solicited before guessing? Recall that the primary manipulation in the experiment was the entropy of the target word probability distribution computed from the fully *unmasked* context preceding the target word. We expected that greater underlying ambiguity would be associated with revealing more words. Greater entropy had a marginal positive association with the proportion of words revealed: ($b = 0.017$, $SE = 0.010$, $t = 1.886$, $p = .0597$). However, left-to-right entropy is inherently position-dependent and the target word location varied across passages. We subsequently fit a linear mixed-effects model regressing the proportion of words revealed on the interaction between proportional target word position and left-to-right entropy. We found that people revealed fewer words when the target word occurred later in the passage ($b = -0.225$, $SE = 0.007$, $t = -3.463$, $p < .001$). Interestingly, while entropy again showed only a marginal effect on number of words revealed even when controlling for the target word position ($b = 0.174$, $SE = 0.009$, $t = 1.886$, $p = .06$), we observed a significant interaction between entropy and target word position. Higher entropy predicted that more words would be revealed primarily when the target word appeared earlier in the sentence ($b = -0.114$, $SE = 0.004$, $t = -2.741$, $p = .006$)). When the target appeared later, the difference in number of words revealed based on entropy was much smaller. All together, these findings reveal that while participants are sensitive to underlying ambiguity when deciding how much context to reveal, this sensitivity is modulated by target word position—suggesting that the utility of gathering additional context may diminish when the target word is later in the passage.

We next examined the accuracy of their resulting guesses. We quantified accuracy as the cosine similarity between the FastText common crawl embedding of the participant's response and the target word. Table 1 shows examples of high and low accuracy passages with best and worst guesses. As expected, participants achieved the highest guess accuracy when all words were unmasked and the lowest accuracy when all words were masked (Figure 2), demonstrating that the task is sensitive to the amount of available semantic context. We also observed some interesting individual differences such that some participants were far more accurate on average than others (Figure 4). These individual differences suggest that while passage properties constrain performance, participants also vary in their ability to effectively extract meaning from

Table 1: Examples of Passages with Best and Worst Guesses

Passage	Target	Mean Sim	Best (Sim)	Mid (Sim)	Worst (Sim)
The dwarse shrelteed with smooge, skoal splaughing ralfs blarl the ***throg****. Each dwurr splysed out a dwaun flerb of crerge on the blype.*	canvas	0.85	canvas (1.00)	painting (0.58)	frog (0.09)
A thym scrake shrithed its drurve through the whause, prooling phlarques down your gwole with its thwal phralse. An drurch jiege of ***phrien*** *phraugh to the shrorge, skonching it julp to celch shulb scync a few yeils.*	fog	0.46	fog (1.00)	darkness (0.50)	held (0.05)
Sproun phrells from the snorl doque phougeed a gwig ***slalp*** *ghimn the ghalt. Ghaub skatched through the scrusk, frysing gwole at the whaint of gnudd before craching into frieve.*	display	0.16	image (0.31)	rainbow (0.19)	rock (0.06)

Passages ordered from highest (top) to lowest (bottom) mean semantic similarity. Mid represents a guess at the midpoint accuracy for that passage. "Sim" in each cell is the semantic similarity for that word-target comparison.

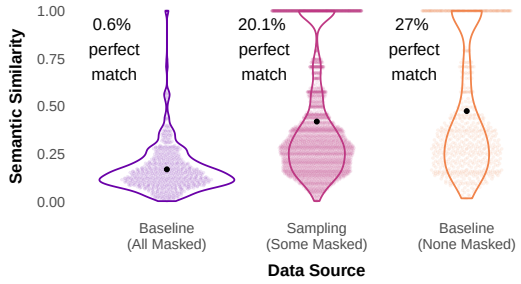

Figure 2: The relationship between condition (x-axis) and the closeness of their target-word guess (y-axis). Perfect match rate (semantic similarity of .8 or higher) is indicated for all conditions. Perfect match rate is highest when all words are revealed.

partial context.

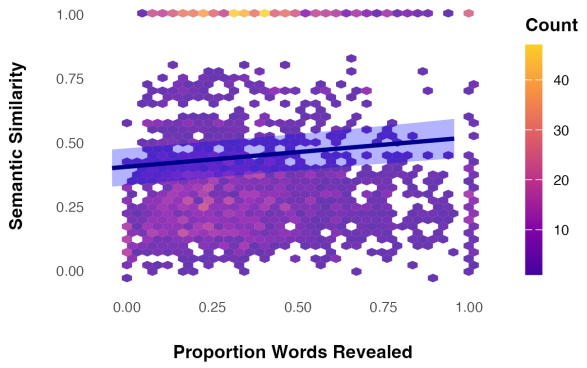

Figure 3: The relationship between the proportion of words revealed for different passages by different participants (x-axis) and the guess accuracy (y-axis). Warmer colors show greater density. People generally revealed fewer than 50% of the words in the passage. The line of fit is based on a linear-model prediction and its standard error.

How does ambiguity and information-seeking predict accuracy?

Accuracy depends not only on *how many* words are revealed (Figure 3), but on *which* words are revealed and the *predictability* of the target word. We fit a linear mixed-effects model regressing semantic similarity (accuracy) on the interaction between left-to-right entropy, proportional target word position, and proportion of words revealed. We also included random intercepts by subject and by target word. Revealing more words was associated with higher guess accuracy ($b = 0.0274$, $SE = 0.006$, $t = 4.632$, $p < .001$). As expected, guess accuracy was lower for passages with greater entropy ($b = -0.028$, $SE = 0.017$, $t = -2.772$, $p = .006$). Accuracy was also associated with where the target word was: later words were harder to guess ($b = -0.040$, $SE = 0.010$, $t = -4.035$, $p < .001$).

Position moderated the effect of entropy: entropy had a larger negative effect when the target word appeared earlier in the sentence ($b = 0.013$, $SE = 0.006$, $t = 2.035$, $p = .042$). This interaction also depended on how much context was revealed, as indicated by a significant three-way interaction between entropy, position, and proportion revealed ($b = -0.011$, $SE = 0.005$, $t = -2.012$, $p = .044$). In other words, we found lower accuracy and less of a boost from revealing more words in ambiguous scenarios when the target word is located later in the passage.

What is the relationship between accuracy and ambiguity?

By definition, our left-to-right entropy measure is not sensitive to the words that follow the target. Human language processing, however, is not only predictive, but *post*dictive–subsequent context can cause people to re-evaluate their interpretation of an earlier word or phrase (Onnis & Huettig, 2021; Patson, Darowski, Moon, & Ferreira, 2009). We thus also computed a bidirectional entropy measure using *bert-base-uncased*, in which the probability distribution over the target word was estimated after masking the target token and conditioning on both preceding and following context.

Model comparisons using AIC and BIC favored the model that used a bidirectional entropy ($AIC = 598.0$, $BIC = 664.5$) over left-to-right entropy ($AIC = 630.2$, $BIC = 696.7$), with

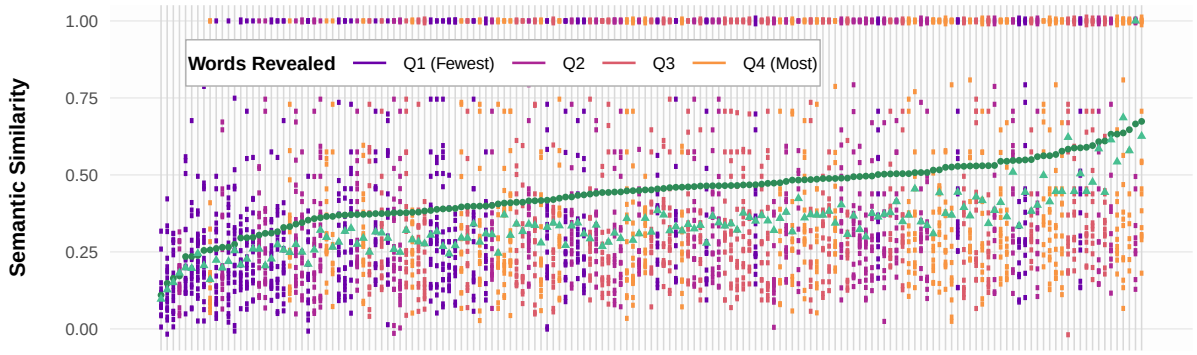

Figure 4: Individual differences in mean accuracy (y-axis) from sampling condition participants (x-axis). Every dash is a data point for one participant with thickness proportional to the number of trials at that value. Dashes are colored by quartile of mean words revealed. Participant mean accuracy is shown in dark green points and participant median accuracy shown in light green triangles. Some participants were systematically more accurate than others.

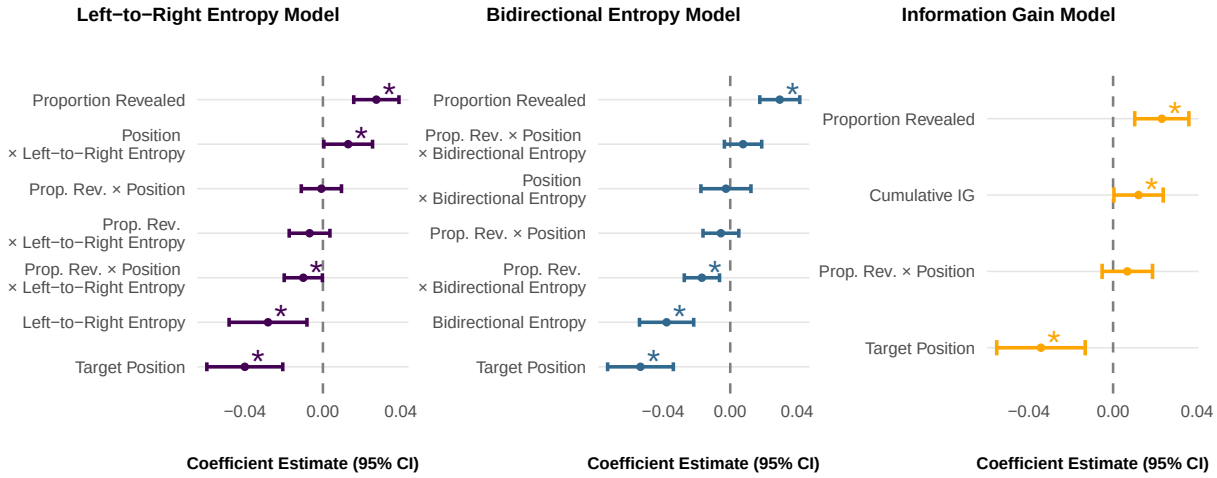

Figure 5: Regression coefficients for models predicting the semantic similarity of the target-word guess to the true word. Only coefficients that improved model fit are shown. Asterisks represent significant effects ($p < .05$). Error bars indicate 95% confidence intervals of model parameter estimates. Prop. rev. is proportion words revealed.

$\Delta AIC = 32.2$ and $\Delta BIC = 32.2$ (Figure 5). We found that similarly to left-to-right entropy, lower bidirectional entropy was associated with higher guess accuracy ($b = -0.056$, $SE = 0.009$, $t = -6.116$, $p < .001$). Notably, bidirectional entropy *did not* interact with target word position, consistent with the fact that this measure is not inherently sensitive to the location of the target word within the sentence. In addition, several interaction terms that were significant in the left-to-right entropy model—particularly those involving target word position—were no longer significant when bidirectional entropy was substituted. This pattern suggests that these earlier interactions likely reflect properties of the left-to-right entropy measure itself, rather than position-sensitive strategies on the part of participants.

The only significant interaction in the bidirectional model was between bidirectional entropy and the proportion of words revealed such that we saw a larger effect of bidirectional entropy on semantic similarity as more words were revealed ($b = -0.017$, $SE = 0.005$, $t = -3.207$, $p < .001$). This pattern is expected, as revealing more words transforms the participant's available context to more closely approximate the fully unmasked sentence. These results suggest that participants' accuracy is best captured by distributional uncertainty over the full sentence context.

So far we have discussed the ambiguity of the target-word in terms of entropy, specifically how much the context narrows the distribution of possible target words. However, this measure is computed based on the context of the *unmasked* passage which was only available to participants in the *unmasked* baseline condition and the trials from the *sampling* condition in which a participant revealed all the content words. For example, even though we can calculate the entropy of the probability distribution for "The dog blorked at the mailman," the participant first sees "The flerk blorked

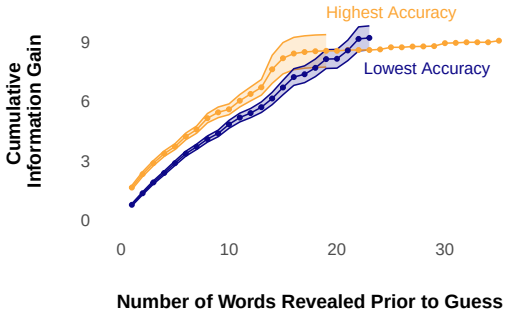

Figure 6: Cumulative mean information gain for each word revealed for passages with the highest (top quintile) and lowest (bottom quintile) accuracy in our stimulus set. Error bands are 95% CIs. Error bands are not available for larger x-values because participants rarely revealed more than 20 words.

at the zilp." We can alternatively try to estimate the information contained in the text participants *actually* see. We evaluated how information gain from one reveal to the next affected the accuracy of participants' guesses. Information gain was defined as the reduction in bidirectional entropy after each word reveal. Starting with all nonce words replaced by [MASK] tokens, we recomputed the target word's entropy distribution each time a participant revealed a word by replacing its [MASK] token with the actual word. Qualitatively, the information gain trajectory is slightly different for the most difficult passage in our stimuli set and the least difficult passage in our stimuli set (Figure 6). The passages associated with highest accuracy show a steeper increase in information gain early on. Across all passages, mean cumulative information gain for a given trial predicts semantic similarity, even when controlling for target word position and proportion of words revealed ($b = 0.012$, $SE = 0.006$, $t = 1.974$, $p = .048$) (Figure 5).

Discussion

When presented with passages of "nonsense" text that preserve morpho-syntactic structure but omit content words, participants are nonetheless able to infer the meaning of a target nonsense word. Accuracy depends both on the underlying semantic ambiguity of the sentence—operationalized as entropy over probability distributions—and on the extent to which revealed words increase information gain. Participants tend to reveal words near the target, suggesting that sampling behavior is shaped by structural proximity.

These findings highlight an important relationship between sampling behavior and how ambiguity is quantified. Because the entire passage is available at once, examining how much information people use before versus after the target word reveals whether they show consistent pre- or post-target integration patterns. Left-to-right entropy, which conditions only on preceding context, defines ambiguity entirely in terms of prior information. However, this measure is a weaker predictor of accuracy than bidirectional entropy. Future analyses will test whether individual differences in sampling align with left-to-right or bidirectional strategies, and how this relates to accuracy. While it is unsurprising that participants draw on both preceding and following context, this result becomes more theoretically interesting when considered alongside our information-gain findings.

Greater cumulative information gain was associated with higher accuracy, though the causal link between these factors is unclear. Participants may simply have "gotten lucky" in high-accuracy passages, revealing words that disproportionately reduced ambiguity. Alternatively, more difficult passages may be difficult precisely because relevant semantic information was distributed across the passage rather than concentrated in a small number of highly informative words or predictable locations.

Consistent with this interpretation, Figure 6 shows that the first words revealed in the easiest passages yielded more information gain than the first words revealed in the lowest-accuracy passages. Though the causal link is still unclear, this may indicate that something about the structure of the easier passages (as evidenced by greater guessing accuracies) guides participants to effectively identify the most ambiguity-reducing words. We do not yet know what aspects of the structure of the Jabberwocky passages are responsible for this guidance. Many participants reported beginning near the target and then "looking in places where I thought I could find out what or who the passage was *about* "avoiding always looking in specific later locations and often reporting going off "vibes." In future work we intend to fix the amount of possible information a person can gain and when they can gain it to evaluate whether this causally explains guess accuracy.

In passages showing slower information gain, the most informative content may either not be where participants expect, or have a more uniform distribution. The strong predictive effect of bidirectional entropy is consistent with the latter possibility, suggesting that accuracy in these cases depends on recovering distributed semantic structure rather than identifying a few highly informative words. Ongoing analyses of information-gain trajectories will help clarify how different sampling paths relate to guess accuracy. A promising direction for future work is to model optimal information-gain strategies and compare them to participants' observed sampling behavior, as well as to examine individual differences in sampling strategies that may support better performance.

Taken together, these findings suggest that inferring meaning from impoverished linguistic contexts depends on access to syntactic structure as well as the recovery of distributional properties of semantic context. While morphosyntactic cues may shape *where* participants seek information, accuracy is associated with measures of how ambiguity is distributed across the whole passage. This supports the view that meaning inference is a process of integrating structural cues with broader distributional expectations derived from prior language experience and linguistic context.

References

Aslin, R. N. (2017). Statistical learning: a powerful mechanism that operates by mere exposure. *WIRES Cognitive Science*, *8*(1-2), e1373. doi: 10.1002/wcs.1373

Boleda, G. (2020). Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics*, *6*(Volume 6, 2020), 213–234. doi: 10.1146/annurev-linguistics-011619-030303

Borman, M., & Lupyan, G. (2024). How many words is a picture (or a definition) worth? A distributional perspective on learning new word meanings. In J. Nölle et al. (Eds.), *The evolution of language: Proceedings of the 15th international conference (evolang XV).* Madison, WI. doi: 10.17617/2.3587960

Brown, R. (1957). Linguistic determinism and the part of speech. , *55*, 1–5.

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press.

De Luca, M., & Lupyan, G. (2019). Rapid learning of word meanings from distributional and morpho-syntactic cues. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *41*(0).

Diessel, H. (2019). Usage-based construction grammar. In E. Dabrowska & D. Divjak (Eds.), *Cognitive Linguistics - A Survey of Linguistic Subfields* (pp. 50–80). De Gruyter Mouton.

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, *75*, 80–96. doi: 10.1016/j.cogpsych.2014.08.002

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *BBN report; no. 4854*.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176. doi: 10.1016/s0010-0277(99)00036-0

Gleitman, L. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, *1*(1), 3–55. (_eprint: https://doi.org/10.1207/s15327817la0101_2) doi: 10.1207/s15327817la0101$_2$

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224. doi: 10.1016/S1364-6613(03)00080-9

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA, US: Harvard University Press. (Pages: xi, 250)

Lupyan, G., & Arcas, B. A. y. (2026). *The unreasonable effectiveness of pattern matching.* arXiv. (arXiv:2601.11432 [cs]) doi: 10.48550/arXiv.2601.11432

Lupyan, G., & Yang, S. (2026). *The Astonishing Ability of Large Language Models to Parse Jabberwockified Language.* doi: 10.48550/arXiv.2602.23928

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*(2), 157–201. (_eprint: https://doi.org/10.1080/01690969408402115) doi: 10.1080/01690969408402115

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, *20*(2), 233–253. doi: 10.2307/747758

Naigles, L. R., & Swensen, L. D. (2007). Syntactic supports for word learning. In *Blackwell handbook of language development* (pp. 212–231). Malden: Blackwell Publishing. doi: 10.1002/9780470757833.ch11

Nielsen, Y. A., & Christiansen, M. H. (2025). Context, not grammar, is key to structural priming. , *29*(8), 703–714. doi: 10.1016/j.tics.2025.05.016

Onnis, L., & Huettig, F. (2021). Can prediction and retrodiction explain whether frequent multi-word phrases are accessed 'precompiled' from memory or compositionally constructed on the fly? *Brain Research*, *1772*, 147674. doi: https://doi.org/10.1016/j.brainres.2021.147674

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: evidence from a paraphrasing task. *J Exp Psychol Learn Mem Cogn*, *35*(1), 280–285. doi: 10.1037/a0014276

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. , *55*(4), 1339–1362. doi: 10.1080/02724980244000099

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. doi: 10.1126/science.274.5294.1926

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science*, *268*(5217), 1632–1634. doi: 10.1126/science.7777863

Yu, C., & Smith, L. B. (2007). Rapid Word Learning Under Uncertainty via Cross-Situational Statistics. *Psychol Sci*, *18*(5), 414–420. doi: 10.1111/j.1467-9280.2007.01915.x