Bridging Narrative Theory and Computational Models: Testing the Transitivity Hypothesis via Parameter-Guided Inference

Yu-Chieh Wang
Student Number: R14142006
Institute of Linguistics
National Taiwan University

Course: Introduction to Syntax
Instructor: Prof. Chia-Rung Lu
Teaching Assistant: Rui-Liang Xu
Date: February 22, 2026

Abstract

The distinction between Foreground (FG) and Background (BG) is a fundamental syntactic interface where clause structure maps onto narrative time. While Hopper and Thompson (1980) proposed that this distinction is governed by a cluster of Transitivity parameters (e.g., Kinesis, Punctuality), the robustness of these parameters as independent grounding cues remains a central question in functional syntax. This study utilizes Large Language Models (LLMs) not as an engineering target, but as a syntactic probe to test the explanatory power of the Transitivity Hypothesis. By injecting explicit parameter constraints into the inference process, this study examines whether syntactic features alone are sufficient to ground narrative events. The ablation study reveals a significant genre-dependent effect: in psychologically ambiguous narrative (*Alice's Adventures in Wonderland*), parameter-guided inference improved grounding accuracy by 11.1%, confirming that transitivity features serve as critical disambiguation cues. However, in high-action folk tale (*Mother Holle*), these parameters induced over-generalization, suggesting that the "high transitivity = foreground" mapping is not a universal rule but a context-sensitive tendency. Furthermore, the observed redundancy of participant features challenges the assumption that parameters such as agentivity are universally discriminative.

Keywords: Transitivity Hypothesis, Foregrounding, Syntax-Semantics Interface, Narrative Structure, Agency

Introduction

A central problem in narrative syntax is the form-function mapping: How do discrete morphosyntactic features—such as Aspect, Mood, and Argument Structure—conspire to signal the discourse function of *Grounding*? In Hopper and Thompson's (1980) seminal work, they famously proposed the Transitivity Hypothesis, arguing that foregrounding is not determined by a single marker (like a tense morpheme) but by the convergence of high-transitivity features across the entire clause. This implies that transitivity is not merely a lexical property of the verb, but a global property of the sentence that facilitates the cognitive processing of events.

While this theory has been influential in functional linguistics, verifying the independent contribution of each parameter is methodologically challenging in naturalistic human processing data. The seamless integration of semantic and syntactic features in human cognition makes it difficult to isolate variables: does a reader identify an event as "foreground" because of its *Kinesis*, or because of the *Volitionality* of the agent? Or is it simply because of world knowledge about what typically constitutes an "event"?

This study introduces a novel methodological approach to address this ambiguity: using Large Language Models (LLMs) as a controlled syntactic simulation. Unlike humans, LLMs do not possess embodied world knowledge; they rely on statistical patterns derived from text. By explicitly manipulating the syntactic inputs provided to the model (a technique termed "parameter-guided inference" in this paper), we can isolate the contribution of specific transitivity parameters. This study poses two linguistically-oriented questions:

1. **RQ1 (Efficacy):** To what extent do explicit transitivity parameters facilitate or interfere with the distinction between foreground and background information in narrative discourse?
2. **RQ2 (Genre Interaction):** Is the efficacy of these parameters universal, or is it dependent on the narrative genre (i.e., syntactic context)?

Significance

This research contributes to functional syntax by stress-testing the Transitivity Hypothesis in a controlled computational environment. If explicit syntactic prompting significantly improves

the model's performance, it provides strong evidence for the computational reality of Hopper and Thompson's categories. It would suggest that these parameters are robust, objective signals that "survive" even in a statistical model, independent of human intuition. Conversely, if certain parameters fail to distinguish foreground from background, it prompts a re-evaluation of their theoretical status in narrative grammar. Thus, the LLM serves here as a mirror to reflect the strengths and limitations of the linguistic theory itself.

Literature Review

Theoretical Framework: The Transitivity Hypothesis

Traditional generative grammar often treats transitivity as a binary property of the verb (transitive vs. intransitive), primarily concerned with the presence or absence of a direct object. However, Hopper and Thompson (1980) revolutionized functional syntax by proposing that transitivity is a scalar, global property of the clause, determined by the convergence of ten morphosyntactic parameters.

The core intuition behind this hypothesis is the concept of energy transfer. A highly transitive clause represents a transfer of energy from a volitional agent to an affected object in a punctual, telic, and realis manner. The authors argue that this transfer of energy is the cognitive basis for distinguishing "events" from "states." The ten parameters (Participants, Kinesis, Aspect, Punctuality, Volitionality, Affirmation, Mode, Agency, Affectedness of Object, Individuation of Object) covary to distinguish the two fundamental dimensions of narrative discourse. High transitivity correlates with foregrounding (the main event line), while low transitivity correlates with backgrounding (supportive descriptions).

Related to the concept of Agency is Dowty's (1991) theory of Proto-Roles. Dowty decomposed the monolithic semantic roles of "Agent" and "Patient" into clusters of contributing properties. For agents, he distinguished between Volitionality (deliberate action) and Sentience (perception). For patients, he emphasized the property of undergoing a change of state.

This framework provides the theoretical granularity needed to explain the redundancy of Participant features observed in this study. In psychological narratives like *Alice's Adventures in Wonderland*, the protagonist often acts as an experiencer (high Sentience, low Volitionality), muddying the Agent signal. Furthermore, many foreground events in this genre are intransitive motion verbs (e.g., *Alice fell*, *she ran*), where the change of state resides in the subject rather than a distinct object. Consequently, the canonical Hopper and Thompson model fails to capture these one-participant events, rendering both Agent and Object parameters nondiscriminatory or even misleading.

The Syntax of Grounding: Foreground vs. Background

To operationalize the concept of grounding for rigorous testing, we must define the syntactic characteristics of its two components.

Foreground (FG) constitutes the skeletal structure of the narrative. These clauses report events that move the narrative time forward. Syntactically, they are typically marked by perfective aspect, dynamic verbs, and high agentivity. They answer the question, *"What happened next?"* (Labov, 1972). In a strict narrative sequence, FG clauses are ordered; changing their order changes the semantic interpretation of the timeline.

Background (BG), conversely, provides the flesh of the narrative. These clauses do not advance the timeline but provide descriptive, durative, or simultaneous information. They amplify, comment on, or set the stage for the main events. Syntactically, BG clauses often feature stative verbs, imperfective or progressive aspect, and non-volitional subjects.

This study acknowledges a potential circularity: Hopper and Thompson's theory aims to explain grounding, yet grounding is often identified by the morphosyntactic features they describe. To mitigate this, Labov's temporal sequentiality test is adopted as the annotation standard. While Hopper and Thompson (1980) themselves draw on Labov's definition of narrative, his test is utilized strictly as an independent functional diagnostic (based on semantic temporality) to generate ground truth labels, ensuring that the target variable is decoupled from the formal syntactic parameters used as predictors in the parameter-guided inference process.

Evolution of Computational Approaches to Narrative Syntax

The intersection of computational modeling and narrative syntax has evolved significantly over the past decades. Early work in narrative schemas (Chambers & Jurafsky, 2008) utilized statistical co-occurrence patterns to learn "chains" of events (e.g., *arrest to convict to jail*) without explicit syntactic supervision. While pioneering, these models treated events as atomic lexical items rather than structured syntactic units, lacking a granular understanding of clause-internal features like aspect or agency.

More recently, the focus has shifted towards discourse parsing. Models trained on RST-DT (Carlson et al., 2003) attempted structural analysis but often struggled with the semantic ambiguity inherent in literary texts. The advent of Large Language Models (LLMs) introduced a new paradigm of syntactic probing (Hewitt & Manning, 2019; Linzen et al., 2016), which demonstrated that models implicitly encode sentence-level syntax in their vector representations.

However, a critical gap remains: while probing studies show that models possess syntactic knowledge, they do not test whether this knowledge is functionally available for higher-level discourse tasks. Most current prompting strategies (e.g., Chain-of-Thought) rely on general semantic reasoning. This study bridges this gap by applying linguistic guidance: distinct from probing, the experiment actively inject Hopper and Thompson's parameters as constraints to test whether explicit syntactic cues can guide a neural network to parse narrative structure more accurately than its baseline probabilistic inference.

Methodology

Methodological Positioning

Before detailing the experimental setup, it is crucial to clarify the epistemological nature of this study. The use of "parameter-guided inference" differs fundamentally from traditional "syntactic probing."

1. Guidance vs. probing Traditional probing aims to discover what linguistic features exist within a model's internal representations (i.e., weights). In contrast, this guidance approach

tests the utility and sufficiency of linguistic cues provided as input. Therefore, the inquiry shifts from assessing the model's latent knowledge of transitivity to evaluating the functional efficacy of explicit constraints in recovering narrative structure.

2. Latent semantics vs. explicit syntax This paper does not claim that LLMs are pure syntactic engines. Modern models possess vast latent semantic knowledge. Therefore, our results must be interpreted as the interaction between the model's priors and our explicit constraints. If adding H&T parameters improves performance, it indicates that these explicit cues provide discriminative power beyond the model's latent semantics.

Data Source and Corpus

To test the hypothesis across different syntactic environments, this study curated a pilot corpus of 111 clause units from two distinct genres.

Alice in Wonderland: Selected to represent a mixed-transitivity environment. Carroll's style features internal monologues and perception verbs. The distinction between an event and a state is often blurred, serving as a stress test for whether syntactic cues can resolve semantic ambiguity.

Mother Holle: Selected as a prototypical high-transitivity environment. This folk tale follows a rigid structure with linear, concrete actions. The mapping between syntax and grounding is expected to be more transparent.

Clause Segmentation Protocol

To ensure consistency, the narratives were segmented into clause units based on rigorous syntactic and semantic criteria designed to isolate distinct predications while preserving aspectual integrity.

The segmentation process adhered to two primary principles. First, the separation principle dictated that coordinated clauses representing distinct temporal steps must be treated as separate units (e.g., *"The Rabbit actually took a watch… [and looked at it]"*). Similarly, main and subordinate clauses were separated when they denoted fundamentally different event types, such as distinguishing physical actions from cognitions. For instance, in the sequence *"[Alice started to*

her feet]… [for it flashed across her mind]", the former is isolated as an action and the latter as a cognition.

Second, the unification principle required that auxiliary and semi-auxiliary constructions remain intact to preserve the clause's aspectual and modal profile. Phase verbs indicating the initiation or continuation of an event were grouped with the main verb (e.g., *"Alice [was beginning to get] very tired"*). Likewise, modals and conative constructions indicating volition or attempt were unified with their complements, as seen in *"She [tried to look] down."* This ensured that the unit of analysis reflected the complete event structure rather than fragmented verbal elements.

Experimental Design

This paper employed a few-shot prompting[1] strategy using the large language model Llama 3.1 (8B) Instruct with temperature set to 0[2]. An instruction-tuned model is specifically chosen rather than a reasoning model. Reasoning models often rely on extensive "Chain of Thought" to infer plot progression from external knowledge, which would confound our variable isolation. Llama 3.1 Instruct behaves more like a compliant syntactic parser, adhering strictly to the provided parameter constraints.

To dissect the specific contributions of the ten transitivity parameters, they were organized into four thematic clusters for the ablation analysis. The first cluster, Agent Properties, encompasses parameters related to the subject's potency, specifically Agency, Volitionality, and the number of Participants. The second, Action Dynamics, groups parameters describing the temporal and physical nature of the event: Kinesis, Aspect, and Punctuality. The third cluster, Object Properties, focuses on the patient of the action through Affectedness and Individuation. Finally, Propositional Status isolates the modal reality of the clause via Mode and Affirmation.

The experimental framework was structured around three comparative conditions designed to isolate the efficacy of syntactic scaffolding. First, the Baseline Condition evaluated the model's inherent grounding judgments based solely on the raw textual context, establishing a performance

[1]Few-shot prompting involves providing the model with a small number of examples in the prompt to guide its performance on a new task.

[2]Temperature is a hyperparameter that controls the randomness of the model's output. Setting it to 0 ensures deterministic and reproducible results.

floor derived from latent semantics. Second, the Full Parameter Condition introduced all ten annotated parameters as explicit constraints, testing the maximal facilitation effect of the Transitivity Hypothesis. Finally, the Ablation Condition systematically withheld specific parameter clusters (e.g., removing all Agent-related tags) to quantify the marginal contribution of each syntactic dimension to the model's decision-making process.

Results

The results reveal a divergence between the two text types, supporting the view that transitivity effects are highly genre-dependent.

Overall Performance Analysis

In *Alice*, the Baseline accuracy was 74.1%. Introducing explicit H&T parameters (Full condition) increased this to 85.2%. This 11.1% improvement suggests that in ambiguous contexts, explicit syntactic features can effectively scaffold the model's decision-making, helping it resolve indeterminacy between states and events. (see Table 1 & Figure 1)

Conversely, in *Mother Holle*, the Baseline was 70.2%, and the Full condition showed no net improvement (0.0% delta). As detailed below, this was not due to a lack of effect, but a shift in error types.

Corpus	Baseline Acc	Full Parameter Acc	Improvement
Alice in Wonderland	0.741	0.852	+11.1%
Mother Holle	0.702	0.702	0.0%

Table 1: Overall Accuracy Comparison. *parameter-guided inference significantly improved performance in the ambiguous narrative but not in the action-heavy text.*

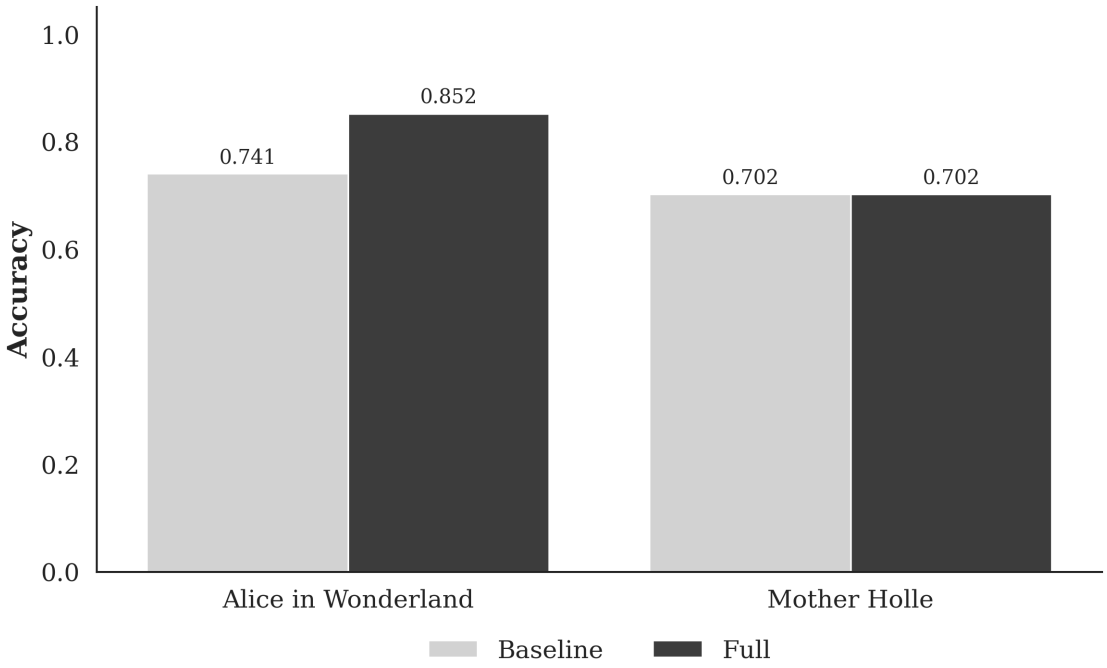

Figure 1: Genre-Dependent Accuracy. *The Full parameter model achieved an 11.1% improvement over the Baseline in the mixed-transitivity narrative, whereas accuracy remained static in the action-heavy narrative.*

The Precision-Recall Trade-off in High-Action Texts[3]

In *Mother Holle*, the Full model exhibited higher recall but significantly lower precision (see Figure 2). The strong transitivity cues caused the model to become over-foregrounding. In a text saturated with action verbs, explicit labels like *Kinesis*: High biased the model to classify almost everything as foreground. This suggests that H&T parameters act as a powerful sensitizing signal, which is beneficial in low-action texts but can be detrimental in high-action environments without careful calibration.

[3]Precision refers to the proportion of predicted FG clauses that are truly FG, while Recall refers to the proportion of actual FG clauses correctly identified. A trade-off implies that improving sensitivity (Recall) comes at the cost of specificity (Precision).

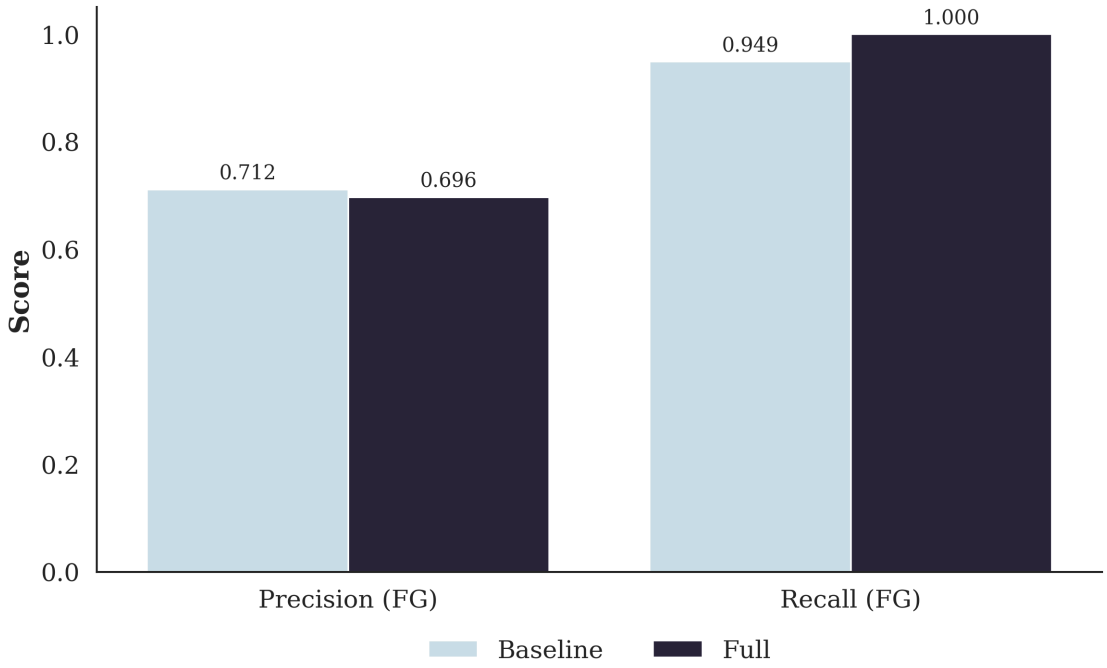

Figure 2: The Precision-Recall Trade-off. *In Mother Holle, the explicit transitivity parameters made the model hyper-sensitive to actions, leading to over-foregrounding errors where background states were misclassified as events.*

The Redundancy of Participant Features

In *Alice's Adventures in Wonderland*, both the No_Agent and No_Object conditions achieved the highest accuracy (87.0%), outperforming the Full model. This convergence suggests a broader phenomenon: in character-centric, single-POV narratives, parameters related to participants (Agency, Volitionality, Affectedness) generally become redundant or noisy.

Since the protagonist (Alice) is the constant subject in most parts of the narrative, agency fails to discriminate between her actions and thoughts. Similarly, the presence of an object is often a poor cue in this genre, as many foreground events are intransitive motion verbs (e.g., *she ran across the field*) while background states can be transitive (e.g., *when she thought it over afterwards*). Consequently, removing these nominal features forces the model to rely on more robust verbal cues like Aspect and Kinesis, thereby improving performance.

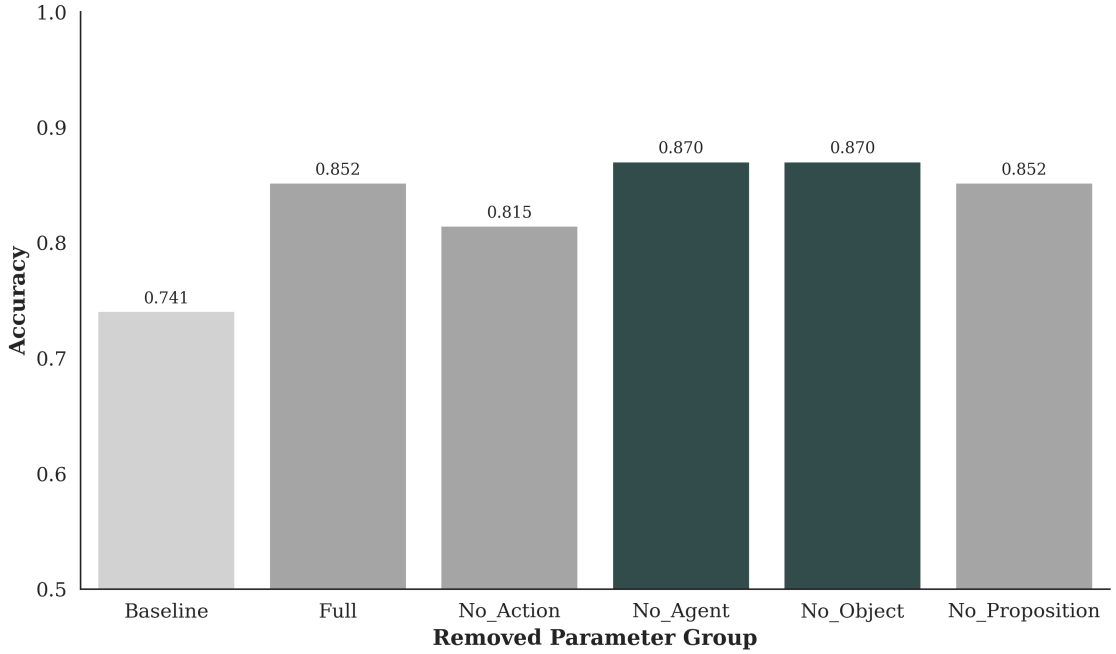

Figure 3: Ablation Study on Alice. *Removing Agent and Object parameters resulted in higher accuracy than the Full model, suggesting that participant-related features may be redundant or distracting in single-POV narratives.*

Discussion

Linguistic Facilitation in Ambiguous Contexts

The improvement in *Alice's Adventures in Wonderland* indicates that syntactic cues are most valuable when semantic boundaries are blurred.

"...once or twice she had peeped into the book her sister was reading"

The Baseline model misclassified this as BG, likely distracted by the static context. The Full model, prompted with *Punctuality: High*, correctly identified the discrete act of *peeping* (FG). This suggests that explicit tags can override local semantic distractions.

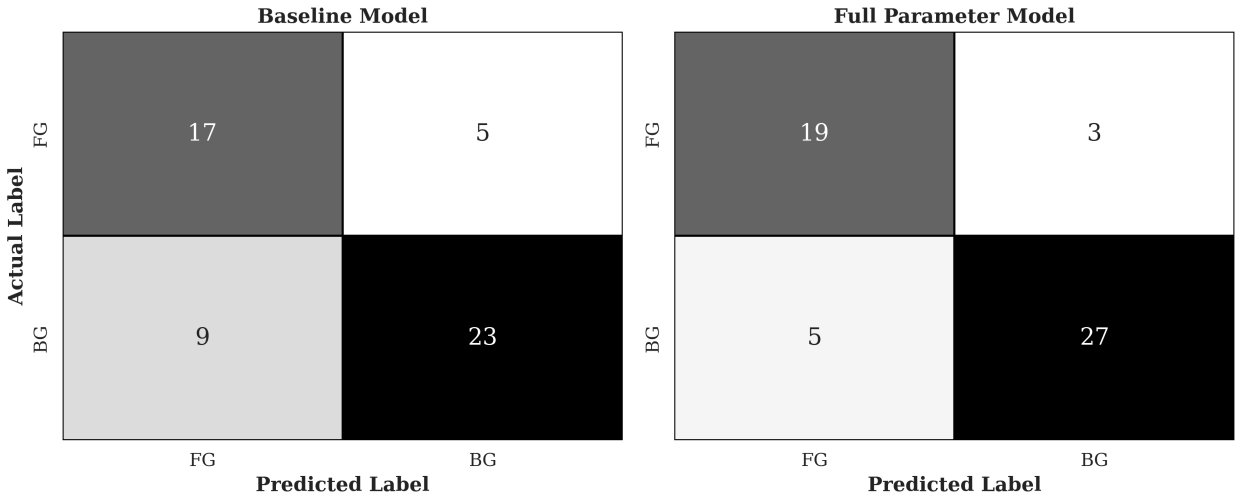

Figure 4: Error Analysis (Alice). *The heatmap compares predictions against ground truth. The Full model (right) shows a reduction in misclassified Foreground events compared to the Baseline (left).*

A Typology of Syntax-Semantics Mismatches

The qualitative analysis reveals that the Transitivity Hypothesis is not a monolithic rule but interacts complexly with semantics. This study observes four recurring patterns of mismatch where explicit guidance either succeeded or failed.

The first type concerns metaphorical high-kinesis states. This occurs when high-energy verbs are used to describe internal states, such as *"the girl summoned up courage"*. The Full model, seeing *Kinesis: High*, systematically over-foregrounded these as Actions. This implies that H&T parameters assume literal energy transfer and lack a mechanism to dampen kinesis in metaphorical contexts.

A second pattern involves cognitive punctuality. In the clause *"for it flashed across her mind,"* the verb *flash* is syntactically punctual. However, semantically, it represents a cognition (Background). The model favored the syntactic cue and misclassified it as FG. This suggests that punctuality must be subordinated to the domain of experience (physical vs. mental).

Third, this paper observed successful scaffolding in cases of archaic semantics. In *"Alice started to her feet"* (meaning "jumped"), the Baseline model likely interpreted "start" as an

aspectual auxiliary. The Full model, guided by *Kinesis: High*, correctly identified it as FG. This represents the ideal use case for scaffolding.

Finally, mismatch arises from presentational constructions. In *"and in the garden stood a tree,"* the locative inversion places focus on the location. Although *stood* is stative, the construction introduces a new entity. The model struggled here, implying that Transitivity parameters do not fully account for information structure (topic-focus).

Theoretical Implications: A Dynamic Model of Parameter Sensitivity

Beyond the specific errors, the findings point to a broader theoretical generalization. This paper proposes a conceptual model centered on the notion of "parameter saturation".

This study observes that a syntactic parameter (such as Kinesis) loses its discriminative power when its frequency of occurrence exceeds a certain threshold. In a high-action genre like *Mother Holle*, Kinesis is ubiquitous. In contrast, in a low-action genre like *Alice's Adventures in Wonderland*, Kinesis is rather rare and thus carries high information density. (see Table 2)

Parameter Cluster	**Low-Transitivity Genre**	**High-Transitivity Genre**
Kinesis / Aspect	High Discriminative Power (High Surprisal)	Saturated / Noisy (Low Surprisal)
Agency / Volitionality	Redundant / Constant (In Single-POV)	Saturated (Default State)
Punctuality	High Discriminative Power	Moderate Discriminative Power
Mode	High Discriminative Power	Constant (Mostly Realis)

Table 2: A Conceptual Model of Parameter Sensitivity. *This table hypothesizes how the "cue validity" of Hopper and Thompson's parameters shifts according to the genre's baseline transitivity.*

It is important to emphasize that this is a conceptual model, not an empirical generalization. While derived from the observations, it serves as a hypothesis generation engine. It predicts that in a genre like academic prose (low transitivity), even weak parameters might become highly discriminative, whereas in sports commentary, almost all parameters might be saturated.

Conclusion

This exploratory study utilizes LLMs to stress-test Hopper and Thompson's Transitivity Hypothesis. By shifting from "probing" internal states to "guiding" inference with explicit constraints, it offers a new methodological lens for functional syntax.

Taken together, the findings support a unified conclusion: transitivity functions as a context-sensitive probabilistic system rather than a fixed diagnostic for grounding. First, regarding the sufficiency of explicit cues, it was demonstrated that in ambiguous contexts, explicit syntactic features are sufficient to guide the model to correct grounding judgments.

In contrast, results regarding the saturation hypothesis indicate that in high-action texts, parameters can become saturated, leading to over-generalization. Finally, the observed redundancy of participant features challenges the universality of agentivity and objecthood as discriminative cues.

It is concluded that the Transitivity Hypothesis should not be operationalized as a rigid morphosyntactic checklist, but as a context-dependent system of cue validity. The findings underscore that the mapping between form (syntax) and function (grounding) is not universal, but is modulated by the specific syntactic profile of the narrative genre.

References

Carlson, L., Marcu, D., & Okurowski, M. E. (2003). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. *Current Directions in Discourse and Dialogue*, 85-112.

Chambers, N., & Jurafsky, D. (2008). Unsupervised learning of narrative event chains. *Proceedings of ACL-08: HLT*, 789-797.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67*(3), 547-619.

Hewitt, J., & Manning, C. D. (2019). A structural probe for finding syntax in word representations. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*.

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language, 56*(2), 251-299.

Labov, W. (1972). *Language in the inner city: Studies in the Black English Vernacular*. University of Pennsylvania Press.

Linzen, T., Dupoux, E., & Goldberg, Y. (2016). Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics, 4*, 521-535.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems, 35*, 24824-24837.

Grimm, J., & Grimm, W. (1812). *Kinder- und Hausmärchen* (Children's and Household Tales). Berlin: Realschulbuchhandlung.

Carroll, L. (1865). *Alice's Adventures in Wonderland*. London: Macmillan.