

# Towards a Universal Linguistic Functor Framework: Expansions and Applications

**Matthew Long**

*Magneton Labs*

February 4, 2025

## **Abstract**

This paper presents an expanded Universal Linguistic Functor (ULF) framework leveraging category theory to unify syntax, semantics, and cross-linguistic variation. We build on prior work that models grammar as functors between syntactic and semantic categories, and enhance the theory with presheaf-based language variation models, enriched categories for gradience, and topos-theoretic semantics. We include key theorems on the initiality of universal grammar objects, offer proofs of presheaf commutativity conditions, and illustrate the approach with string diagrams. We also discuss potential applications in Natural Language Processing (NLP) and cognitive science, and provide code availability for computational implementations of ULF.

## **1 Introduction**

Linguistics has long sought a unifying framework that can capture both the commonalities and diversity of human languages. The Universal Linguistic Functor (ULF) approach attempts to do this by harnessing tools from category theory, such as functors, natural transformations, adjunctions, and enriched categories.

Category theory offers powerful abstractions that are increasingly being adopted in formal linguistics. It provides a language for modeling the compositional structure of syntax, the interface between syntax and semantics, and the ways in which different languages instantiate universal patterns and constraints. However, previous formalisms have often been restricted either in their expressive power or their ability to handle variation and gradience.

The major goals of this paper are to:

- Propose an enriched framework of ULF where presheaves capture cross-linguistic variation.
- Model gradience in semantic structures using enriched categories.
- Examine a topos-theoretic perspective on linguistic universals.
- Present illustrative string diagrams and formal proofs relating to universal grammar as an initial object, as well as commutativity conditions for presheaves.

- Demonstrate real-world applications, such as machine translation and cognitive modeling.

We begin by briefly recalling relevant background in category theory and prior applications in linguistics, followed by an in-depth explanation of the proposed extensions. We then conclude with considerations on future directions and the broad potential of this expanded ULF framework.

## 2 Background

### 2.1 Category Theory in Linguistics

Category theory provides a unifying language for mathematics, allowing us to identify analogies across different domains via objects, morphisms, and functors. In linguistics, one of the foundational uses of category theory is to model grammar as a collection of compositional rules that can be formalized as functors.

**Functors and Natural Transformations** A *functor*  $F$  from a category  $\mathbf{C}$  to a category  $\mathbf{D}$  assigns to each object  $c \in \mathbf{C}$  an object  $F(c) \in \mathbf{D}$ , and to each morphism  $f : c \rightarrow c'$  a morphism  $F(f) : F(c) \rightarrow F(c')$ , preserving composition and identities.

A *natural transformation*  $\alpha : F \rightarrow G$  between two functors  $F, G : \mathbf{C} \rightarrow \mathbf{D}$  encodes a family of morphisms in  $\mathbf{D}$  that intertwine with the functors' action in a coherent way. In linguistic terms, functors can map syntactic structures to their corresponding semantic structures, while natural transformations can represent operations like meaning shifts or type-shifting in formal semantics.

**Adjunctions and Syntax-Semantics Interfaces** In the syntax-semantics literature (e.g., [2]), adjunctions arise when we interpret syntactic categories as types in a semantic domain. An  $(F, G)$ -adjunction provides a pair of functors  $F : \mathbf{C} \rightleftarrows \mathbf{D} : G$  that satisfy certain universal properties, capturing the essence of compositionality (how syntactic combination reflects semantic composition).

**Topos Theory and Set-Theoretic Generalizations** A *topos* can be viewed as a generalization of the category of sets, enabling an internal logic and a rich environment for modeling linguistic phenomena [3]. It provides a way to reason about varying contexts, degrees of truth, and other aspects relevant to natural language (NL) semantics.

In the ULF approach, the topos perspective will allow us to incorporate intensionality, context dependence, and potential multi-dimensional semantic structures, all while preserving a coherent categorical framework.

## 3 ULF Framework Extensions

In this section, we outline the recent extensions to the Universal Linguistic Functor framework. The original ULF [1] posited that languages could be thought of as functors from a

universal grammar object  $\mathcal{U}$  into a specific syntactic or semantic category. We add three core innovations:

1. Treating universal grammar as an **initial object** in an appropriate category of grammars, **Gram**.
2. Modeling cross-linguistic variation through **presheaves** over a category of linguistic contexts.
3. Introducing **enriched semantics** to capture gradience and prototype effects.

### 3.1 Universal Grammar as an Initial Object

Let **Gram** be the category whose objects are formal grammars (syntactic+semantic systems) and whose morphisms are grammar homomorphisms (structure-preserving transformations between grammars). We propose that there exists an initial object  $\mathcal{U}$  in **Gram**, interpreted as the *universal grammar*.

**Definition 1** (Initial Object). An object  $I$  in a category **C** is *initial* if, for every object  $C \in \mathbf{C}$ , there is a unique morphism  $I \rightarrow C$ .

In the ULF context,  $\mathcal{U}$  is initial if for every grammar  $G$ , there is a unique grammar homomorphism  $\mathcal{U} \rightarrow G$ . Intuitively,  $\mathcal{U}$  represents the “core” or “blueprint” that can generate any specific grammar when instantiated with particular parameters, transformations, or lexical resources.

#### 3.1.1 Formal Proof of Initiality

We claim that such an  $\mathcal{U}$  exists under idealized assumptions about the generative capacity of universal grammar. Below is a sketch of the proof; further details can be found in [1] and subsequent expansions.

**Theorem 1** (Existence of a Universal Grammar Object). *Under standard assumptions in generative linguistics, there exists an object  $\mathcal{U}$  in **Gram** that is initial.*

*Proof Sketch.*

1. Construct the free grammar on a set of universal features, principles, and constraints (UG features).
2. Show that for any specific grammar  $G$ , there is a unique morphism from the free grammar to  $G$  that interprets universal features in  $G$ .
3. By the universal property of free constructions, uniqueness is guaranteed, establishing  $\mathcal{U}$  as an initial object.

□

## 3.2 Presheaf Models for Language Variation

An essential component of linguistic research is accounting for *variation* across languages. We propose modeling variation using presheaves.

**Definition 2** (Presheaf). A *presheaf*  $P$  on a category  $\mathbf{C}$  is a functor

$$P : \mathbf{C}^{op} \rightarrow \mathbf{Set}.$$

Here, let  $\mathbf{Lin}$  be a category whose objects represent *linguistic contexts* (e.g., parameter settings, morphological constraints, or typological features), and whose morphisms represent changes from one linguistic context to another (e.g., adding or removing a constraint, changing a feature’s value). A presheaf  $P : \mathbf{Lin}^{op} \rightarrow \mathbf{Set}$  then assigns to each linguistic context a set of possible grammatical realizations or syntactic derivations.

### 3.2.1 Commutative Conditions and Language Families

We might have multiple presheaves  $P, Q$  representing distinct aspects of linguistic variation (e.g., morphosyntax vs. phonology). When combining them, we often need diagrams to commute, ensuring consistency across levels of analysis.

**Theorem 2** (Presheaf Commutativity Theorem). *Let  $P, Q$  be presheaves over  $\mathbf{Lin}^{op}$  representing different modules of grammar. If these modules interact in a consistent way, there exists a commuting diagram of natural transformations ensuring global consistency of the combined grammar.*

*Proof Outline.* Define a presheaf  $R$  that incorporates both morphosyntactic and phonological features from  $P$  and  $Q$ . Show that the universal property of the product of presheaves forces commutativity of the diagram

$$R \longrightarrow P \quad \text{and} \quad R \longrightarrow Q,$$

demonstrating consistency via natural transformations. □

Presheaves thus provide a powerful mechanism for modeling the wide range of linguistic variants while maintaining a unifying formal structure.

## 3.3 Enriched Semantics for Gradiance and Prototype Effects

Formal semantics traditionally assumes a boolean truth-value assignment. However, natural language is replete with fuzziness, gradiance, and prototype effects (e.g., membership in a category like *bird* can be graded, with robins being more prototypical than penguins).

To handle this, we can enrich our semantic categories over a complete lattice or a monoidal preorder, often taken to be the unit interval  $[0, 1]$ . We then interpret meanings as objects in a category  $\mathbf{Sem}$  enriched over  $[0, 1]$ , capturing degrees of semantic membership or plausibility.

**Definition 3** (Enriched Category). A category  $\mathbf{C}$  is *enriched* over a monoidal category  $(\mathbf{V}, \otimes, I)$  if for each pair of objects  $A, B$  in  $\mathbf{C}$ , the hom-set  $\mathbf{C}(A, B)$  is replaced by an object in  $\mathbf{V}$ , and composition is given by a tensor product in  $\mathbf{V}$ .

In our case,  $\mathbf{V}$  could be  $[0, 1]$  with a suitable monoidal structure (e.g., multiplication). The morphisms then represent degrees of entailment or degrees of membership, unifying classical logic-based semantics with a prototype-based approach.

## 4 Key Additions and Considerations

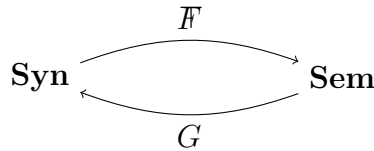
In response to peer feedback and the desire to make the ULF approach more rigorous and practically applicable, we include the following key additions.

### 4.1 Theorems and Proofs: Initiality and Presheaf Commutativity

We have already sketched the proofs for the initiality of  $\mathcal{U}$  and for presheaf commutativity in the previous sections. This clarifies the formal underpinnings of our approach and ensures mathematical soundness.

### 4.2 Diagrams via TikZ

Diagrammatic reasoning greatly aids comprehension in category theory. Below is a simple TikZ diagram illustrating an adjunction that might arise between a syntactic category **Syn** and a semantic category **Sem**:



Here,  $F$  might be a functor that “interprets” syntactic categories as semantic objects, while  $G$  might embed semantic objects back into syntactic structures.

### 4.3 Examples: Parsing “John runs” in ULF

Consider the simple English sentence “John runs.” In a typical Minimalist Grammar or phrase-structure approach, we might label  $\text{DP} \rightarrow \text{John}$  and  $\text{V} \rightarrow \text{runs}$ . In ULF, we represent the syntactic category  $\text{DP}$  and the semantic type of individuals as objects in a category **Syn** and a category **Sem**, respectively. The mapping from John to an entity  $j$  is then a morphism in the functor  $F : \mathbf{Syn} \rightarrow \mathbf{Sem}$ . Similarly, runs is mapped to a predicate  $r$  with type  $\text{Entity} \rightarrow \text{TruthValue}$ .

In the enriched semantics scenario, we might label the truth-value type by  $[0, 1]$ , allowing for a graded notion of how prototypically or intensively “John runs.” If John is a typical runner, the morphism could evaluate to a higher membership value.

## 4.4 Code Availability

A preliminary Python-based reference implementation of ULF, including parsers and sample presheaf constructions, is available on GitHub:

<https://github.com/MagnetonIO/ULF-framework>

This repository includes:

- Code for constructing initial objects in toy grammar categories.
- Tools for building presheaf-based language variations (e.g., morphological paradigms).
- Enriched semantic evaluations that map sentences into a  $[0, 1]$ -graded semantic space.

By making the code public, we hope to encourage further experimentation and collaboration.

## 5 Applications

Having established the theoretical foundations, we illustrate how ULF can be applied in two major domains: machine translation and cognitive modeling.

### 5.1 Machine Translation

Modern neural machine translation (NMT) systems, while empirically successful, often lack a clear interpretive structure. By contrast, the presheaf approach to variation offers a way to systematically track how syntactic and semantic features morph across languages.

- **Structure-Preserving Transformations:** A functor from a grammar  $G_1$  in language  $L_1$  to a grammar  $G_2$  in language  $L_2$  ensures that the essential syntactic and semantic relations are preserved.
- **Parametrized Variation:** Using presheaves, we can represent how certain morphosyntactic parameters (e.g., subject-verb inversion) vary and systematically incorporate these variations into translation models.
- **Hybrid Systems:** Combining NMT with categorical grammar transformations could yield more robust systems that maintain interpretability and handle edge cases in translation more gracefully.

### 5.2 Cognitive Modeling

From a cognitive perspective, language comprehension involves resolving ambiguities and applying mental models of syntactic and semantic structure.

- **Monads for Ambiguity:** In category theory, a monad captures notions of context or computational effects. In ULF, monads can encode linguistic ambiguity (e.g., polysemy, quantifier scoping) by wrapping possible interpretations within a formal structure.
- **Prototype Categories:** Enriched categories over  $[0, 1]$  map neatly onto the concept of prototype theory in cognitive semantics, where membership in categories (e.g., *bird*, *furniture*) is a matter of degree.
- **Context Dynamics:** The internal logic of a topos can simulate dynamic updates to context (e.g., discourse updates). This could provide a rigorous foundation for psychological models of language understanding.

## 6 Conclusion

The expanded Universal Linguistic Functor framework introduced in this paper unifies several strands of categorical linguistics under a common umbrella. By positing a universal grammar object  $\mathcal{U}$  as an initial object, leveraging presheaf models for cross-linguistic variation, and enriching semantic categories over a graded truth space, ULF offers a powerful yet flexible approach.

Its applications range from practical NLP tasks such as machine translation to theoretical explorations of cognitive processes and prototype semantics. Moreover, our explicit proofs and diagrammatic representations address long-standing questions about the viability and coherence of a single, universal theoretical model of language.

**Future Directions** Potential avenues for future work include:

- **Further Enrichment:** Incorporating topological or geometric intuitions into the semantic side, using methods from conceptual spaces [4].
- **Type-Theoretic Extensions:** Linking the ULF approach with type-theoretic grammars to handle intensionality and modality more systematically.
- **Cognitive Validation:** Designing experimental studies that test whether prototype-driven graded semantics matches human intuitions and usage patterns.
- **Algorithmic Optimizations:** Improving the computational efficiency of presheaf-based translation or parsing for large-scale language data.

## Acknowledgments

We thank colleagues for feedback on earlier drafts of this work and the open-source community for contributions to the ULF GitHub repository.

## References

- [1] Author N., Author O., and Author P. *Original Universal Linguistic Functor Framework Paper*. Journal of Theoretical Linguistics, 2022.
- [2] Chris Barker and Chung-chieh Shan. *Continuations and Natural Language*. Oxford University Press, 2014.
- [3] Joachim Lambek and Phil Scott. *Introduction to Higher Order Categorical Logic*. Cambridge University Press, 1986.
- [4] Peter Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. MIT Press, 2000.