Construction-Grammar Informed Parameter Efficient Fine-Tuning for Language Models

Prasanth Yadla
Independent Researcher
USA
pyadla2@alumni.ncsu.edu

Abstract

Large language models excel at statistical pattern recognition but may lack explicit understanding of constructional form-meaning correspondences that characterize human grammatical competence. This paper presents Construction-Aware LoRA (CA-LoRA), a parameter-efficient fine-tuning method that incorporates constructional templates through specialized loss functions and targeted parameter updates. We focus on five major English construction types: ditransitive, caused-motion, resultative, way-construction, and conative. Evaluation on BLiMP, CoLA, and SyntaxGym shows selective improvements: frequent patterns like ditransitive and caused-motion show improvements of approximately 3.3 and 3.5 percentage points respectively, while semi-productive constructions show minimal benefits (1.2 points). Overall performance improves by 2.4 percentage points on BLiMP and 2.4 points on SyntaxGym, while maintaining competitive performance on general NLP tasks. Our approach requires only 1.72% of trainable parameters and reduces training time by 67% compared to full fine-tuning. This work demonstrates that explicit constructional knowledge can be selectively integrated into neural language models, with effectiveness dependent on construction frequency and structural regularity.

1 Introduction

Construction Grammar fundamentally reconceptualizes linguistic knowledge as a network of form-meaning mappings called constructions, ranging from morphemes to abstract syntactic patterns (Goldberg, 1995; Fillmore et al., 1988). This theoretical framework proposes that speakers acquire grammatical competence through learning conventionalized associations between linguistic forms and their semantic interpretations, treating all linguistic knowledge as constructions of varying complexity and schematicity.

The constructionist approach offers several theoretical advantages for computational language modeling. Unlike generative approaches that separate lexicon from grammar, Construction Grammar provides a unified framework for both compositional and non-compositional linguistic phenomena. Constructions explicitly encode form-meaning correspondences, making them ideal candidates for integration into neural architectures that traditionally rely on implicit pattern recognition. The usage-based orientation of Construction Grammar aligns naturally with statistical learning paradigms underlying modern language models.

Despite these theoretical advantages, mainstream natural language processing has largely overlooked Construction Grammar insights. Current transformer-based models learn linguistic patterns through statistical exposure to large corpora but lack explicit representation of constructional knowledge (Brown et al., 2020; Devlin et al., 2018). This creates a disconnect between theoretical understanding of grammatical competence and practical implementation in language technology.

Recent work has demonstrated the potential for integrating linguistic theory into neural language models through parameter-efficient fine-tuning approaches (Hu et al., 2021). These methods enable targeted adaptation of large models while preserving general capabilities and maintaining computational efficiency. However, previous approaches have focused primarily on syntactic constraints rather than constructional form-meaning mappings.

This paper addresses this gap by introducing Construction-Aware LoRA (CA-LoRA), a parameter-efficient fine-tuning approach that explicitly integrates Construction Grammar principles into transformer-based language models. Our method treats constructions as learnable templates that specify both formal patterns and semantic interpretations, enabling models to develop explicit constructional competence.

Proceedings of the Second International Workshop on Construction Grammars and NLP, pages 34–40

Benchmark	RoBERTa-large	Standard LoRA	CA-LoRA
BLiMP Overall	76.8	77.4	**79.2**
Argument Structure	73.2	74.1	**76.4**
Filler-Gap Dependencies	74.6	75.3	**77.1**
Island Effects	69.7	70.2	**71.8**
CoLA (MCC)	0.618	0.631	**0.649**
SyntaxGym	69.3	70.1	**71.7**

Table 1: Performance on linguistic evaluation benchmarks

We make four primary contributions to construction-aware language modeling. First, we develop a framework for representing major English constructions as explicit templates that can be integrated into neural training processes. Second, we present CA-LoRA, a parameter-efficient method that embeds constructional knowledge into language models through targeted parameter updates and specialized loss functions. Third, we demonstrate that constructional fine-tuning improves performance on linguistic benchmarks that test understanding of argument structure and form-meaning correspondences. Finally, we show that our approach maintains computational efficiency while achieving these linguistic competence gains.

2 Construction Grammar Framework

2.1 Theoretical Foundations

Construction Grammar emerged from recognition that traditional linguistic theories inadequately account for the pervasive role of learned form-meaning pairings in language use (Fillmore et al., 1988; Goldberg, 1995). The theory posits that linguistic knowledge consists entirely of constructions—conventionalized associations between form and meaning that speakers acquire through exposure to usage events.

Constructions exhibit several key properties that distinguish them from traditional grammatical rules. They represent holistic form-meaning mappings that cannot be derived purely through compositional processes from their component parts. They exist at multiple levels of abstraction, from fully specified lexical items to highly schematic syntactic patterns. They contribute meaning independently of their lexical fillers, explaining coercion phenomena where verbs acquire constructional semantics not present in their basic meanings.

The ditransitive construction exemplifies these principles. The pattern *[Subject Verb Object1 Object2]* carries inherent transfer semantics regardless of the specific verb involved. This explains how *"She baked him a cake"* receives a transfer interpretation despite *bake* not being inherently a transfer verb. The construction contributes transfer meaning through coercion, demonstrating how form-meaning mappings operate independently of lexical semantics.

2.2 Argument Structure Constructions

Argument structure constructions represent a well-studied domain within Construction Grammar, encompassing basic clause-level patterns that specify participant roles and event semantics (Goldberg, 1995). These constructions demonstrate clear form-meaning correspondences that extend beyond what can be predicted from lexical properties alone.

Our framework focuses on five major English argument structure constructions that exhibit systematic form-meaning relationships:

Ditransitive Construction: *[NP-Agent V NP-Recipient NP-Theme]* $\leftrightarrow$ TRANSFER(agent, theme, recipient)

This pattern encodes successful transfer events, as in *"She gave him the book"* and *"He taught her Spanish"*. The construction contributes transfer semantics that may be absent from the verb's core meaning.

Caused-Motion Construction: *[NP-Agent V NP-Theme PP-Goal]* $\leftrightarrow$ CAUSE-MOVE(agent, theme, goal)

This construction expresses caused motion events, exemplified by *"He kicked the ball into the net"* and *"She pushed the cart down the aisle"*. The pattern can coerce non-motion verbs into motion interpretations.

Resultative Construction: *[NP-Agent V NP-Patient XP-Result]* $\leftrightarrow$ CAUSE-BECOME(agent, patient, result-state)

Resultative patterns encode causation of result states, as in *"They painted the house red"* and *"He wiped the table clean"*. The construction provides result-state meaning that extends basic action semantics.

Way-Construction: *[NP-Agent V Poss way PP-Path]* $\leftrightarrow$ MANNER-MOTION(agent, manner, path)

This semi-productive pattern expresses manner of motion, illustrated by *"She danced her way across the stage"* and *"He fought his way through the crowd"*. The construction creates motion interpretations for non-motion verbs.

Conative Construction: *[NP-Agent V at NP-Target]* $\leftrightarrow$ ATTEMPTED-ACTION(agent, target)

The conative alternation expresses attempted rather than successful action, contrasting *"She shot the deer"* (successful) with *"She shot at the deer"* (attempted). The prepositional marking contributes aspectual meaning.

2.3 Constructional Templates

We formalize constructions as structured templates that specify both formal constraints and semantic interpretations. Each construction C is represented as:

$$C = \langle \Phi, \Sigma, \Theta \rangle \tag{1}$$

where Φ defines the formal template including syntactic categories and linear order, Σ specifies the semantic frame with participant roles and event structure, and Θ represents frequency-based weighting derived from corpus observations.

For the ditransitive construction, this yields:

$$C_{\text{ditrans}} = \langle [\text{NP}_{\text{agent}} \text{ V } \text{NP}_{\text{recipient}} \text{ NP}_{\text{theme}}], \tag{2}$$

$$\text{TRANSFER}(\text{agent}, \text{theme}, \text{recipient}), \tag{3}$$

$$\theta_{\text{transfer}} = 0.34\rangle \tag{4}$$

This representation captures both the syntactic pattern and associated semantic frame while incorporating usage frequency information that influences constructional processing priorities.

3 Construction-Aware LoRA

3.1 Parameter-Efficient Constructional Integration

We develop Construction-Aware LoRA (CA-LoRA), a parameter-efficient fine-tuning method that integrates constructional templates into transformer-based language models. CA-LoRA operates on the principle that constructional competence can be achieved through targeted parameter updates that encode form-meaning correspondences without disrupting general language capabilities.

The approach extends standard LoRA (Hu et al., 2021) by introducing construction-specific adaptation matrices that capture the statistical dependencies underlying each constructional pattern. For each construction C and transformer weight matrix $W_0 \in \mathbb{R}^{d \times k}$, we define construction-specific low-rank adaptations:

$$W_C = W_0 + \sum_{i=1}^{n} \alpha_i \Delta W_i^C \tag{5}$$

where $\Delta W_i^C = A_i^C B_i^C$ represents the low-rank adaptation for construction C, with $A_i^C \in \mathbb{R}^{d \times r}$ and $B_i^C \in \mathbb{R}^{r \times k}$ where $r \ll \min(d, k)$. The scaling factors α_i control the relative influence of each constructional adaptation.

This architecture allows multiple constructions to be simultaneously encoded through separate LoRA modules, enabling the model to access different constructional patterns during inference. The parameter-efficient nature ensures that constructional knowledge can be integrated without the computational overhead of full model retraining.

3.2 Construction-Guided Training Objective

We develop a specialized training objective that encourages models to learn constructional form-meaning correspondences through targeted supervision. The objective combines standard language modeling with construction-specific learning signals derived from our template representations.

The total loss function integrates multiple components:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{LM}} + \beta \sum_{C \in \mathcal{C}} \mathcal{L}_C \tag{6}$$

where $\mathcal{L}_{\text{LM}}$ represents the standard language modeling loss, $\mathcal{C}$ denotes the construction inventory, $\mathcal{L}_C$ provides construction-specific supervision for pattern C, and β is a weighting factor that controls the relative importance of the construction losses.

Each construction-specific loss component encourages appropriate usage of the corresponding pattern:

$$\mathcal{L}_C = -\mathbb{E}_{s\sim D_C}\big[\log P(s \mid C)\big] \tag{7}$$

$$+ \lambda\,\mathbb{E}_{s\sim D_{\neg C}}\big[\max(0, \log P(s \mid C) - \tau)\big] \tag{8}$$

where D_C contains sentences that instantiate construction C, $D_{\neg C}$ contains sentences that violate constructional constraints, and τ represents a margin parameter that discourages high probability assignment to malformed patterns.

This formulation rewards models for recognizing and generating appropriate constructional patterns while penalizing violations of form-meaning correspondences. The approach enables direct supervision of constructional competence without requiring extensive manual annotation.

3.3 Multi-Construction Processing

Real language use involves interactions between multiple constructions, requiring models to handle constructional composition and selection. Our CA-LoRA framework addresses this through dynamic construction activation mechanisms that determine which patterns are relevant for specific inputs.

We implement construction selection through attention-based gating that computes relevance scores for each construction given input context:

$$w_C(x) = \text{softmax}(\text{MLP}_C(\text{pooled}(x))) \tag{9}$$

where x represents input embeddings and MLP_C provides construction-specific scoring. The final representation combines weighted contributions from all constructions:

$$h_{\text{final}} = \sum_{C\in\mathcal{C}} w_C(x) \cdot h_C(x) \tag{10}$$

This approach enables flexible constructional processing that captures the probabilistic and gradient nature of constructional activation in human language use, where multiple patterns can simultaneously influence interpretation and production.

4 Experimental Setup

4.1 Model Architecture and Training Data

We implement CA-LoRA using RoBERTa-large and GPT-2 medium as base architectures, representing both encoder-only and decoder-only transformer variants. LoRA adaptations are applied to attention projection matrices and feed-forward layers with rank $r = 16$ for attention components and $r = 32$ for feed-forward networks, based on preliminary experiments balancing expressivity with efficiency.

Training data consists of carefully selected subsets from BookCorpus (Zhu et al., 2015) and OpenWebText (Gokaslan and Cohen, 2019), totaling approximately 12GB of diverse text across multiple domains and registers. This corpus selection ensures exposure to varied constructional patterns while maintaining manageable computational requirements for parameter-efficient training.

The training process involves constructional pattern identification through template matching against our five target construction types. We use constituency parsing and semantic role labeling to identify potential constructional instantiations, then apply template matching to extract positive and negative training examples for each construction type.

Hyperparameter optimization explores construction loss weights $\beta \in \{0.1, 0.3, 0.5\}$ and margin parameters $\tau \in \{0.5, 1.0, 2.0\}$ using validation performance on a held-out subset of training data. Learning rates are tested across $\{1e{-}4, 3e{-}4, 5e{-}4\}$ with batch sizes of 16 to balance training stability with memory constraints.

4.2 Baseline Comparisons

We compare CA-LoRA against several baseline approaches that represent different methods for incorporating linguistic knowledge into language models. Standard LoRA fine-tuning provides a direct comparison, using the same training data and parameter-efficient architecture without constructional supervision.

Full fine-tuning baselines demonstrate the computational advantages of parameter-efficient approaches while providing upper bounds on potential performance gains from increased model plasticity. These models are trained on identical data with the same constructional objectives but update all model parameters.

Prompt-based approaches, while not presented here, test whether constructional knowledge can be effectively communicated through natural language descriptions rather than parameter updates, providing insights into the necessity of direct architectural integration for constructional competence.

Task	RoBERTa-large	Standard LoRA	CA-LoRA
GLUE Average	84.2	84.6	84.7
Reading Comprehension (SQuAD 2.0)	81.3	81.7	81.5
Sentiment Analysis (SST-2)	91.8	92.1	92.3
Natural Language Inference (MNLI)	86.4	86.8	86.6
Semantic Similarity (STS-B)	88.1	88.4	88.5

Table 2: Performance on general NLP tasks. Differences between Standard LoRA and CA-LoRA.

Method	Trainable Params	Training Time	Memory Usage	Performance Gain
Full Fine-tuning	355M (100%)	38.7 hours	26.8 GB	+2.1%
Standard LoRA	1.2M (0.34%)	12.4 hours	14.3 GB	+0.6%
CA-LoRA	6.1M (1.72%)	12.8 hours	14.7 GB	+2.4%

Table 3: Computational efficiency comparison for RoBERTa-large. Performance gain measured on linguistic benchmarks relative to base model.

5 Results

5.1 Linguistic Benchmark Performance

Table 1 presents results on established linguistic evaluation benchmarks, demonstrating consistent improvements from constructional fine-tuning across tasks that test grammatical competence.

CA-LoRA achieves meaningful improvements across linguistic benchmarks, with particularly notable gains of 3.2 percentage points on argument structure tasks and 2.4 points on overall BLiMP (Warstadt et al., 2020) performance. These results including CoLA (Warstadt et al., 2019) and SyntaxGym (Gauthier et al., 2020) demonstrate that explicit constructional training enhances performance on phenomena that require understanding of form-meaning correspondences and argument role relationships.

The improvements are most pronounced on tasks that directly test constructional competence, such as argument structure alternations and role assignment. This suggests that CA-LoRA successfully integrates constructional knowledge in ways that transfer to related linguistic phenomena.

5.2 Construction-Specific Analysis

Table 4 evaluates performance on tasks specifically designed to test each target construction type, providing detailed analysis of constructional learning effectiveness.

CA-LoRA demonstrates variable improvements across construction types, with gains ranging from 1.2 percentage points for way-constructions to 3.5 points for caused-motion patterns. The ditransitive construction shows a 3.3 point improvement (71.8 $\rightarrow$ 75.1), while resultative constructions show modest gains of 1.7 points (66.4 $\rightarrow$ 68.1). These results indicate that explicit constructional supervision enhances competence for well-defined form-meaning mappings, though benefits vary considerably by construction type and frequency. The performance pattern reflects both constructional frequency and structural complexity in the training data. Frequent, clearly-defined patterns like caused-motion (3.5 point improvement) and ditransitive (3.3 points) show substantial gains, while semi-productive constructions like way-constructions (1.2 points) and resultatives (1.7 points) show minimal improvement. This suggests that template-based approaches work best for constructions with clear syntactic patterns and consistent semantic roles, but struggle with more creative or contextually-dependent patterns that rely heavily on pragmatic inference.

5.3 General NLP Task Performance

Table 2 demonstrates that constructional fine-tuning maintains competitive performance on standard NLP benchmarks while achieving specialized linguistic competence.

The results show that CA-LoRA maintains performance within typical variation margins across standard benchmarks, indicating that constructional specialization does not compromise general language understanding capabilities. This supports the viability of our parameter-efficient approach for practical applications.

5.4 Computational Efficiency

Table 3 compares training costs across different approaches, highlighting the efficiency advantages of parameter-efficient constructional learning. CA-

Construction Type	Baseline	CA-LoRA
Ditransitive	71.8 ± 2.1	**75.1 ± 1.9**
Caused-Motion	69.1 ± 3.4	**72.6 ± 2.7**
Resultative	66.4 ± 2.8	68.1 ± 3.2
Way-Construction	60.2 ± 4.1	61.4 ± 3.8
Conative	64.1 ± 2.6	**67.2 ± 2.9**
Overall Average	66.3 ± 1.8	**69.0 ± 1.6**

Table 4: Construction-specific performance (accuracy %). Results averaged over 5 random seeds.

LoRA achieves superior performance gains while maintaining reasonable efficiency compared to full fine-tuning, requiring only 1.72% of trainable parameters and 67% less training time than full fine-tuning. While CA-LoRA uses approximately 5 times more parameters than standard LoRA, it remains highly parameter-efficient relative to full model retraining. The modest increase in memory usage (2.8%) reflects constructional processing overhead without fundamentally altering the parameter-efficient paradigm. The trade-off between CA-LoRA and standard LoRA involves exchanging some parameter efficiency for improved performance on linguistically-oriented tasks.

6 Analysis and Discussion

6.1 Constructional Learning Patterns

Analysis of learned parameters reveals that CA-LoRA develops distinct representational patterns for different construction types. Attention weight visualization shows increased focus on constructionally relevant features, such as recipient arguments in ditransitive constructions and result states in resultative patterns.

Probing experiments using linear classifiers demonstrate that constructional information becomes more linearly separable in CA-LoRA representations compared to baseline models. This indicates that parameter-efficient adaptation successfully embeds constructional distinctions into model representations in ways that support systematic processing.

6.2 Form-Meaning Correspondence

Qualitative analysis of model outputs demonstrates enhanced sensitivity to constructional form-meaning correspondences. CA-LoRA models show improved ability to distinguish between well-formed constructional instantiations and violations, such as correctly rejecting *“She donated him money” while accepting “She donated money to him.”

The models also demonstrate better handling of constructional coercion phenomena, correctly interpreting sentences like “She sneezed the napkin off the table” where the caused-motion construction provides motion semantics absent from the verb’s core meaning.

6.3 Limitations and Future Directions

Current CA-LoRA implementation focuses on English argument structure constructions and requires language-specific template definitions. Extending to other languages will need development of language-appropriate constructional inventories and consideration of typological differences in form-meaning mapping strategies.

The template-based approach may miss subtle constructional distinctions that require deeper semantic or pragmatic analysis. Future work should investigate integration of richer semantic representations and world knowledge to capture the full complexity of constructional phenomena.

Scale limitations prevent evaluation on the largest current language models, though our parameter-efficient approach should facilitate application to models with hundreds of billions of parameters. Future research should investigate how constructional learning scales with model size and training data volume.

7 Conclusion

This work demonstrates that Construction Grammar principles can be effectively integrated into neural language models through parameter-efficient fine-tuning, achieving meaningful improvements in constructional competence while maintaining computational efficiency and general language capabilities. Our Construction-Aware LoRA approach provides a practical framework for incorporating theoretical linguistic insights into modern NLP systems.

The key findings establish that explicit constructional templates can enhance language model performance on tasks requiring understanding of form-meaning correspondences and argument structure relationships. Parameter-efficient methods enable integration of constructional knowledge without the computational overhead of full model retraining. Constructional fine-tuning improves linguistic competence while preserving general language understanding capabilities across diverse tasks.

Future research should explore extension to broader constructional inventories, multilingual constructional learning, and integration with larger-scale language models. Investigation of constructional learning in very large models could reveal whether explicit constructional guidance remains beneficial at scale or whether implicit statistical learning eventually captures these patterns automatically.

This work represents a step toward bridging theoretical linguistics and computational language modeling, demonstrating that Construction Grammar insights can inform and improve neural language processing systems. By explicitly encoding form-meaning correspondences, we open possibilities for more linguistically sophisticated and interpretable language models that better align with human grammatical competence.

References

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, and 1 others. 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Charles J. Fillmore, Paul Kay, and Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 64(3):501–538.

Jon Gauthier, Jennifer Hu, Ethan Wilcox, Peng Qian, and Roger Levy. 2020. Syntaxgym: An online platform for targeted evaluation of language models. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 70–75.

Aaron Gokaslan and Vanya Cohen. 2019. Openwebtext corpus. Accessed: 2019-09-05.

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

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Alex Warstadt, Alicia Parrish, Haokun Liu, Anhad Mohananey, Wei Peng, Sheng-Fu Wang, and Samuel R Bowman. 2020. Blimp: The benchmark of linguistic minimal pairs for english. *Transactions of the Association for Computational Linguistics*, 8:377–392.

Alex Warstadt, Amanpreet Singh, and Samuel R Bowman. 2019. Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics*, 7:625–641.

Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 19–27.