Annotating English Verb-Argument Structure via Usage-Based Analogy

Allen Minchun Hsiao
University of Colorado Boulder
Min-Chun.Hsiao@colorado.edu

Laura A. Michaelis
University of Colorado Boulder
Laura.Michaelis@colorado.edu

Abstract

This paper introduces a usage-based framework that models argument structure annotation as nearest-neighbor classification over verb–argument structure (VAS) embeddings. Instead of parsing sentences separately, the model aligns new tokens with previously observed constructions in an embedding space derived from semi-automatic corpus annotations. Pilot studies show that cosine similarity captures both form and meaning, that nearest-neighbor classification generalizes to dative alternation verbs, and that accuracy in locative alternation depends on the corpus source of exemplars. These results suggest that analogical classification is shaped by both structural similarity and corpus alignment, highlighting key considerations for scalable, construction-based annotation of new sentence inputs.

1 Introduction

Verbs provide a crucial interface between syntax and semantics, typically determining both the nature of the event or action described by a clause and the number and type of participants the clause contains—a configuration known as **argument structure**. For example, the verb *give* denotes an act of possession transfer and therefore requires three arguments: an `AGENT` (*Paul*), a `RECIPIENT` (*me*), and a `THEME` (*a book*). These may be realized syntactically as the Double Object construction, as in *Paul gave me a book*. At the same time, proponents of construction-based syntax have observed that verbs may be mismatched to their syntactic contexts in ways that alter the meaning and valence of the verb. For example, while creation verbs like *paint* and *draw* do not intrinsically express acts of transfer, they can be used to implicate actual or intended transfer in sentences like *I drew her a picture*. Such examples suggest that argument-structure patterns themselves can convey event structures traditionally attributed to verbs alone and, in turn, may influence the verb's meaning and selectional properties (Goldberg, 1995; Michaelis, 2004).

To obtain argument structures from natural language, most NLP systems rely on automatic Semantic Role Labeling (SRL) and constituency parsing. SRL identifies argument spans and assigns semantic roles (Màrquez et al., 2008; Gildea and Jurafsky, 2002), while constituency parsing extracts hierarchical phrase structures (Marcus et al., 1993). However, these two components are often modeled separately, leading to cascading errors: syntactic misparses can degrade SRL accuracy. Recent approaches integrate syntactic information into neural SRL models (Strubell et al., 2018; Zhou et al., 2020; Fei et al., 2021), and BERT-based architectures frame SRL as span classification without explicit syntactic features (Shi and Lin, 2019). Meanwhile, high-accuracy constituency parsers like the Berkeley Neural Parser (Kitaev and Klein, 2018) continue to be widely used in such pipelines.

Yet despite their strong performance, these systems are optimized for SRL as a classification task and maintain a verb-centered view of argument structure. Even models that integrate syntax typically do so only to improve SRL performance. Moreover, most SRL datasets, such as PropBank and VerbNet, rely on verb-specific argument frames, which limit generalization across constructions. Arguments introduced by constructions—rather than verbs—are often overlooked. For instance, in "*kick the ball into the room*," the directional PP *into the room* is typically treated as an adjunct, despite fulfilling a core semantic role (`Goal`) in a Caused-Motion construction. As a result, these models fall short of capturing the full range of construction-based argument structures observed in natural language.

This paper introduces a usage-based alternative that models argument structure through analogical matching against previously observed VAS patterns. Instead of parsing a sentence and mapping its elements via fixed templates, our model compares the sentence to a library of VAS exemplars and selects the best match in embedding space. This nearest-neighbor approach treats argument structure as a

Proceedings of the Second International Workshop on Construction Grammars and NLP, pages 158–164
September 24, 2025,

product of linguistic experience, and contextual inference, rather than static verb valency.

The remainder of this paper is organized as follows: Section 2 reviews Construction Grammar-based approaches to argument structure annotation; Section 3 introduces our model; Section 4 details the methodology, including data curation and annotation; Section 5 reports pilot studies; and Section 6 concludes with discussion and future directions.

2 Related Work

Resources grounded in Construction Grammar (CxG) aim to annotate argument structures that arise not only from lexical valency but also from constructional licensing. Kyle and Sung (2023) introduce the first argument structure construction (ASC) treebank, manually annotating verb-argument structures following CxG principles. While valuable, the treebank covers only a small set of constructions, limiting its generalizability.

Perek and Patten (2019) explore the empirical identification of constructions using syntactic n-grams extracted from the British National Corpus. They cluster these "treelets" by distributional similarity and manually select a linguistically meaningful subset. This work lays a data-driven foundation for construction identification in English but requires extensive manual intervention and remains a work in progress.

Computational frameworks like Fluid Construction Grammar (FCG) (Beuls and Van Eecke, 2023) offer a cognitively motivated architecture for representing argument structure via learned form-meaning pairings. FCG supports both parsing and production, modeling the dynamic invocation of constructions during language use. However, despite its expressive power, FCG relies on hand-engineered constructions and operates primarily in simulation environments or controlled domains. It lacks scalable interfaces with large corpora or pretrained models. As a result, while FCG demonstrates the theoretical utility of construction-based approaches, it is not yet suitable for automatically annotating verb-argument structures in real-world corpora. Because of this limitation, we continue to seek scalable, data-driven alternatives.

3 Our Model

This project introduces a usage-based, analogy-driven framework for annotating VAS in natural language. Drawing on exemplar-based approaches to grammar (Bybee, 2013), the model treats argument structure annotation as a **nearest-neighbor retrieval task** grounded in **analogical matching** (Gentner, 1983, 2010). Given a sentence containing a target verb, it selects the most likely structure by comparing the verb's contextual embedding to a set of previously observed, type-level verb-argument structure embeddings. In essence, it asks: "Which known pattern does this usage most closely resemble?"

Rather than assuming each verb is tied to a fixed valency frame, the model is built on the idea that argument structures generalize across verbs. For example, the structure associated with *give*, as in *She gave him a book*, may serve as an exemplar for annotating *bequeath*, as in *She bequeathed him a book*. Such generalizations are achieved through analogical matching: the model ranks known structures by their cosine similarity to the target usage, offering plausible annotations even for novel or infrequent verbs.

Each verb-argument structure in the model is represented as an embedding derived from actual corpus attestations. These type-level embeddings are stored and compared against token-level embeddings extracted from new input sentences. The top-ranked structures—those most similar in both form and meaning—are returned as candidate annotations. This ranked list supports both automatic labeling and human-in-the-loop annotation, functioning as an assistant that provides interpretable and transparent suggestions.

4 Methodology

This section outlines the steps used to develop our model, from data selection and annotation to the construction of a semantic space of verb-argument structures (VAS). We divide the methodology into three components: (1) data curation, (2) verb-argument structure annotation, and (3) construction of the VAS space.

4.1 Data Curation

Corpus Selection. We use a subset of the BabyLM Project Gutenberg corpus (Warstadt et al., 2023), which contains written English texts from books in the public domain. Our goal is not exhaustive annotation but the development of a representative VAS space from high-quality language data. We sample 51,411 sentences for analysis.

Data Filtering. To focus on clause-level predicates, we filter the data using spaCy (Honnibal et al., 2020). We retain only sentences where the main verb is the syntactic `ROOT` and has a nominal subject (`NSUBJ`). This reduces the dataset to 30,139 sentences. We further exclude malformed or fragmentary sentences, yielding a final dataset of 23,396 well-formed sentences.

4.2 Verb-Argument Structure Annotation

Initial semantic-syntactic auto-annotation. We begin by automatically annotating each verb-argument structure with syntactic phrase types and semantic roles. Semantic roles are assigned using SemParse (Gung, 2020), which maps predicates to VerbNet classes and extracts PropBank-style arguments; these are then converted to VerbNet roles using the mappings in Kipper-Schuler et al. (2008). Syntactic categories are obtained from the Berkeley Neural Parser (Kitaev and Klein, 2018), from which we extract the highest syntactic projection of each argument. These initial annotations are used as a base for further revision.

Construction-Based Revision. Following principles in Construction Grammar (Goldberg, 2006; Michaelis, 2012), we revise the initial annotations to reflect arguments introduced not only by the verb's lexical valency but also by larger constructions. For example, in the sentence *She kicked the ball into the room*, the directional phrase *into the room* is labeled as a `Goal` argument—not as an adjunct—because it is licensed by the Caused-Motion construction rather than by the verb *kick* alone. These construction-based revisions ensure that the final annotations more accurately capture the full range of argument structure patterns observed in natural usage.

4.3 Constructing the VAS Space

Embedding Extraction. Each verb-argument structure instance is represented as a contextualized embedding extracted from BERT (Devlin et al., 2019), using layer 7 (Chronis and Erk, 2020). Embeddings are grouped by VAS type and averaged to yield a type-level embedding.

Linguistic Experience Space. The resulting VAS space serves as a structured repository of linguistic experience. Each type-level embedding encodes the distributional and constructional properties of a verb-argument structure as observed in corpus data. Given a new sentence, the model retrieves the most similar structure in the space using cosine similarity. This usage-based approach reflects how speakers interpret novel utterances by analogizing to familiar patterns encountered in prior language use.

5 Pilot Studies

This section reports three pilot studies. The first examines what cosine similarity between verb embeddings captures. The second and third test how well the model can classify unseen verb tokens using a small set of precomputed structure embeddings.

5.1 Pilot 1: What Does Cosine Similarity Capture?

The first pilot study examines what cosine similarity between verb token embeddings captures, since this similarity metric underlies our method for selecting candidate structures. Understanding what it reflects—**surface form** (i.e., syntactic realization), **relational meaning** (i.e., argument structure roles), or **both**—is essential for evaluating the validity of our analogical classification approach.

We test this using the verb *bequeath*, which alternates between the **Prepositional Dative** and **Double Object** constructions. For each form, we compare its embedding to three minimally altered sentences, varying only the verb while holding the surrounding context constant. This isolates the verb's syntactic and semantic contribution as the source of variation in cosine similarity.

Double Object variant:

The widow bequeathed the church her property.
<$\mathtt{NP1}_{\mathtt{Agent}}$, $\mathtt{NP2}_{\mathtt{Recipient}}$, $\mathtt{NP3}_{\mathtt{Theme}}$>

- *The widow gave the church her property.* (form and meaning) → cosine similarity: 0.7541
- *The widow gave her property to the church.* (meaning only) → cosine similarity: 0.7243
- *The widow considered the church her property.* (form only) → cosine similarity: 0.5481

Prepositional Dative variant:

The widow bequeathed her car to the church.
<$\mathtt{NP1}_{\mathtt{Agent}}$, $\mathtt{NP2}_{\mathtt{Theme}}$, $\mathtt{PP}_{\mathtt{Recipient}}$>

- *The widow gave her car to the church.* (form and meaning) → cosine similarity: 0.7753
- *The widow gave the church her car.* (meaning only) → cosine similarity: 0.7119

- *The widow drove her car to the church.* (form only) $\rightarrow$ cosine similarity: 0.5962

In both constructions, the sentence sharing both form and meaning with the target had the highest similarity, while form-only matches scored lowest. Meaning-only matches consistently ranked in between. This pattern indicates that **cosine similarity in our embedding space captures both syntactic and semantic similarity, with a stronger bias toward meaning**. These results support the use of cosine similarity for analogical classification and align with the usage-based view that novel utterances are understood through semantic alignment with familiar constructions.

5.2 Pilot 2: Nearest-Neighbor Classification Accuracy on Dative Alternation Verbs

To test our model's classification accuracy, we evaluated whether unseen verb tokens could be correctly assigned argument structure labels by comparing their embeddings to five precomputed VAS embeddings of *give*. We tested both **within-verb generalization**—predicting new *give* tokens drawn from COCA—and **cross-verb generalization**—predicting an **unseen** verb *bequeath* from COCA.

5.2.1 Experimental Setup

We manually annotated five distinct VAS types for *give* from the Gutenberg corpus (see Appendix A.1). Each structure combined specific semantic roles and phrase types and served as a prediction template.

The test set included 40 sentences from COCA: 20 with *give* and 20 with *bequeath*, each evenly split between the Double Object and Prepositional Dative constructions. Each verb token was embedded using BERT (layer 7) and matched to the most similar *give* structure embedding based on cosine similarity.

5.2.2 Results and Discussion

Table 1 shows that the model achieved high accuracy across both within-verb and cross-verb conditions. For *give*, which appeared in the training corpus (Gutenberg), the model correctly predicted all 10 Double Object tokens and 9 out of 10 Prepositional Datives from the test set (COCA). For *bequeath*, which was unseen during training, the model correctly classified all 20 tokens.

These results suggest that our model can generalize both to new uses of familiar verbs and to entirely new verbs that share similar constructions. The success of cross-verb classification, especially for a rare verb like *bequeath*, indicates that the precomputed structure embeddings of *give* encode transferable, construction-level information. This supports our central hypothesis: **verb-argument structure annotation can be modeled as a nearest-neighbor classification task in a semantically structured space**.

Verb	Argument Structure	Accuracy
give	Double Object	100% (10/10)
	Prepositional Dative	90% (9/10)
bequeath	Double Object	100% (10/10)
	Prepositional Dative	100% (10/10)

Table 1: Nearest-neighbor classification accuracy for *give* and *bequeath* using *give* structure embeddings.

5.3 Pilot 3: Nearest-Neighbor Classification Accuracy on Locative Alternation Verbs

We next tested our model on the verb *spray*, which alternates between the Caused-Motion (CM) construction (e.g., *He sprayed the paint onto the wall*) and the Theme-Applicative (TA) construction (e.g., *He sprayed the wall with paint*). This alternation provides an ideal case study because it involves two competing argument structure frames that are both frequent and semantically transparent, yet distinct in terms of syntactic realization.

Experimental setup. We assembled 100 tokens of *spray*, balanced between 50 CM and 50 TA tokens. Of these, 20 CM and 30 TA tokens were drawn directly from COCA, and each was paired with an altered counterpart in the alternate construction (e.g., a TA token such as *Pinocchio sprays Puss with water* was paired with its CM variant *Pinocchio sprays water to Puss*). This procedure yielded a balanced dataset where every naturally attested token was matched with a constructed counterpart, ensuring equal representation of both constructions.

Two VAS type embeddings served as classifiers. For the CM frame, we used the *put* pattern (*He put the money into the pocket*), averaged from 66 CM *put* tokens in Gutenberg. For the TA frame, we used the *cover* pattern (*He covered his beard with his hands*), averaged from 50 TA tokens in COCA.[1] Both *put* and *cover* serve as prototypical

[1]We did not use the TA *cover* pattern from Gutenberg due

exemplars of their respective constructions, making them suitable analogical anchors for classification.

Results with Gutenberg *put*. When paired against the COCA *cover* embedding, the Gutenberg *put* embedding produced an accuracy of 0.690 and a macro F1 of 0.662 (Table 4). Predictions were heavily skewed toward the TA category: of 100 spray tokens, 79 were classified as TA, yielding an F1 of 0.760 for TA but only 0.563 for CM. The full confusion matrix is shown in Table 2.

Gold ARGST	Predicted: TA	Predicted: CM
TA	**49**	1
CM	**30**	20

Table 2: Confusion matrix for *spray* prediction using Gutenberg *put* (CM) vs. COCA *cover* (TA).

This imbalance suggested that the skew might not be due to structural similarity alone, but instead to corpus mismatch: both *spray* and the TA source (*cover*) came from COCA, while the CM source (*put*) came from Gutenberg. To test this speculation, we repeated the experiment using a CM *put* embedding drawn from COCA rather than Gutenberg.

Results with COCA *put*. Substituting 50 CM *put* tokens from COCA yielded stronger performance: accuracy rose to 0.860 and macro F1 also reached 0.860 (Table 4). Predictions were more balanced, with F1 scores of 0.865 for TA and 0.854 for CM. The corresponding confusion matrix is shown in Table 3.

Gold ARGST	Predicted: TA	Predicted: CM
TA	**45**	5
CM	9	**41**

Table 3: Confusion matrix for *spray* prediction using COCA *put* (CM) vs. COCA *cover* (TA).

Discussion. Taken together, the results summarized in Table 4 suggest that analogical classification may be affected by the corpus where the source patterns are drawn. When sources and targets were drawn from different corpora (Gutenberg vs. COCA), predictions skewed heavily toward the COCA source. When both sources were drawn from COCA, predictions became more balanced and overall accuracy improved. Although these findings are preliminary, they indicate that corpus alignment could interact with structural similarity in shaping analogical predictions. Future work will test this more systematically across additional verbs, constructions, and corpora.

to its limited size ($n = 13$).

	CM-*put* (Gut.) + TA-*cover* (COCA)	CM-*put* (COCA) + TA-*cover* (COCA)
Accuracy	0.690	**0.860**
Macro F1	0.662	**0.860**
F1 (TA)	0.760	**0.865**
F1 (CM)	0.563	**0.854**

Table 4: Summary of nearest-neighbor prediction performance for *spray*.

6 Conclusion

We presented a usage-based model that operationalizes argument structure annotation as a nearest-neighbor classification task over verb–argument structure (VAS) embeddings. By aligning new sentences with previously encountered constructions in a multidimensional embedding space, the model reflects how speakers interpret novel expressions—not by parsing syntax and semantics separately, but by recognizing patterns grounded in prior linguistic experience.

Our pilot studies illustrate both the promise and the challenges of this approach. Pilot 1 showed that cosine similarity captures differences in both form and meaning, with a stronger bias toward meaning. Pilot 2 demonstrated that nearest-neighbor classification can model argument structure in dative alternation verbs, and together with Pilot 3, showed that a single VAS type embedding can support accurate prediction across verbs. Pilot 3 further scaled up to locative alternation verbs and revealed that accuracy also depends on corpus source: predictions were more accurate when sources and targets came from the same corpus. These findings suggest that analogical classification is shaped not only by structural similarity but also by corpus alignment, pointing to key considerations for future large-scale work.

The framework is designed for continuous refinement: new structure types and attestations can be added over time, allowing it to evolve alongside linguistic theory and empirical data. This scalability supports interpretable annotation and underscores the value of high-quality, construction-based analysis—even in the era of large embedding models.

References

Katrien Beuls and Paul Van Eecke. 2023. Fluid construction grammar: State of the art and future outlook. In *Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+NLP, GURT/SyntaxFest 2023)*, pages 41–50, Washington, D.C. Association for Computational Linguistics.

Joan L. Bybee. 2013. Usage-based theory and exemplar representations of constructions. In *The Oxford Handbook of Construction Grammar*, volume 1. Oxford University Press.

Gabriella Chronis and Katrin Erk. 2020. When is a bishop not like a rook? when it's like a rabbi! multi-prototype BERT embeddings for estimating semantic relationships. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 227–244, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Hao Fei, Shengqiong Wu, Yafeng Ren, Fei Li, and Donghong Ji. 2021. Better combine them together! integrating syntactic constituency and dependency representations for semantic role labeling. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 549–559, Online. Association for Computational Linguistics.

Dedre Gentner. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2):155–170.

Dedre Gentner. 2010. Bootstrapping the mind: analogical processes and symbol systems. *Cognitive Science*, 34(5):752–775.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.

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

Adele E. Goldberg. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press, Oxford, UK.

James Gung. 2020. SemParse, VerbNet Parser.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Karin Kipper-Schuler, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of english verbs. *Language Resources and Evaluation*, 42:21–40.

Nikita Kitaev and Dan Klein. 2018. Constituency parsing with a self-attentive encoder. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2676–2686, Melbourne, Australia. Association for Computational Linguistics.

Kristopher Kyle and Hakyung Sung. 2023. An argument structure construction treebank. In *Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+NLP, GURT/SyntaxFest 2023)*, pages 51–62, Washington, D.C. Association for Computational Linguistics.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.

Lluís Màrquez, Xavier Carreras, Kenneth C. Litkowski, and Suzanne Stevenson. 2008. Special issue introduction: Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34(2):145–159.

Laura A. Michaelis. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics*, 15(1):1–67.

Laura A. Michaelis. 2012. Making the case for construction grammar. In Hans C. Boas and Ivan A. Sag, editors, *Sign-Based Construction Grammar*, pages 31–69. CSLI Publications, Stanford, CA.

Florent Perek and Amanda Patten. 2019. Towards an inventory of english argument structure constructions: Empirical assessment of the n-gram approach. *Constructions and Frames*, 11(2):250–282.

Peng Shi and Jimmy Lin. 2019. Simple bert models for relation extraction and semantic role labeling. *arXiv preprint arXiv:1904.05255*. Work in progress.

Emma Strubell, Patrick Verga, Daniel Andor, David Weiss, and Andrew McCallum. 2018. Linguistically-informed self-attention for semantic role labeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5027–5038, Brussels, Belgium. Association for Computational Linguistics.

Alex Warstadt, Leshem Choshen, Aaron Mueller, Adina Williams, Ethan Wilcox, and Chengxu Zhuang. 2023. Call for papers – the babylm challenge: Sample-efficient pretraining on a developmentally plausible corpus.

Junru Zhou, Zuchao Li, and Hai Zhao. 2020. Parsing all: Syntax and semantics, dependencies and spans. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4438–4449, Online. Association for Computational Linguistics.

Appendix

A.1 Verb-Argument Structure Types for *give*

Below are the five verb-argument structure types of *give* used in the classification model, each accompanied by an illustrative sentence from the Gutenberg corpus:

1. $<\texttt{NP1}_{\texttt{Agent}},\ \texttt{NP2}_{\texttt{Recipient}},\ \texttt{NP3}_{\texttt{Theme}}>$
   *They **gave** him an opportunity of speaking more, and therefore he thought himself better than the rest.*

2. $<\texttt{NP1}_{\texttt{Agent}},\ \texttt{NP2}_{\texttt{Recipient}},\ \texttt{S[QUE+]}_{\texttt{Theme}}>$
   *Simonetta **gave** her mother what was indispensable for household expenses and managed the rest herself.*

3. $<\texttt{NP1}_{\texttt{Agent}},\ \texttt{NP2}_{\texttt{Theme}},\ \texttt{PP}_{\texttt{Recipient}}>$
   *He should have **given** the deer to the woman.*

4. $<\texttt{NP1}_{\texttt{Agent}},\ \texttt{NP2}_{\texttt{Theme}},\ \texttt{PP}_{\texttt{Beneficiary}}>$
   *With the same humanity which they had shown in the case of Jogues, they **gave** a generous ransom for him, supplied him with clothing, kept him until his strength was in some degree recruited, and then placed him on board a vessel bound for Rochelle.*

5. $<\texttt{NP1}_{\texttt{Agent}},\ \texttt{NP2}_{\texttt{Theme}}>$
   *[...] the magician **gives** the order for preparations.*