Progressive Coherence Building in Relational Understanding: A Hierarchical Cognitive Model

Zhen Wang[1], **Yu Wang**[1] **& Wen Zhao**[2]
[1]School of Software and Microelectronics, Peking University, Beijing, China
[2]National Engineering Research Center for Software Engineering, Peking University, Beijing, China

Abstract

Human readers derive coherent relational understanding from information distributed across a document through a progressive, hierarchical process: local semantic associations are established first and then integrated into global coherence. How the mind achieves such robust integration while limiting combinatorial explosion and error propagation remains an open question. We propose progressive coherence building as a key computational principle, in which complex inference is constrained through sequentially layered, verifiable representations. Guided by this hypothesis, we introduce a Hierarchical Cognitive Model (HCM) for cross-context relational learning. HCM instantiates three cognitively motivated stages: (1) lexical-semantic grounding via prompt-based semantic priming; (2) local proposition formation through intra-triple attention that enforces local semantic consistency; and (3) global coherence optimization using prior-guided axial attention to integrate evidence across the narrative. This staged design stabilizes intermediate representations and mitigates error propagation. Experiments on CDR, GDA, and DocRED demonstrate state-of-the-art performance. Ablation results reveal cumulative degradation when higher stages are removed, empirically supporting the proposed hierarchical processing account.

Keywords: Progressive Coherence; Hierarchical Cognitive Model; Error Propagation; Computational Cognitive Science; Relational Understanding

Introduction

How does the human mind comprehend complex relationships scattered across a narrative or discourse? Consider reading a biography where multiple facts about a person—birthplace, education, career moves, and collaborations—are distributed across paragraphs. Successful understanding requires more than extracting isolated facts; it demands the *progressive building of coherence*: integrating local pieces of information into a structured, globally consistent mental model of how entities relate to one another. This cognitive process is hierarchical and staged: we first grasp simple, local connections (e.g., "Person A graduated from University X") before using them as anchors to infer more complex, distributed relations (e.g., "Person A and Person B likely collaborated because they overlapped at the same institution").

As shown in Fig.1, current computational models of Document-level Relation Extraction (DocRE) aim to automate this kind of understanding. However, most current approaches lack an explicit commitment to modeling this *progressive, hierarchical nature* of coherence building. Graph-based models (Wang et al., 2020) reason over pre-defined structures but often treat inference as a single-step process. Sequence- and table-based models (Yao et al., 2019; N. Zhang et al., n.d.), while powerful, typically perform "flat" classification on entity pairs, attempting to predict complex relations directly from textual cues without simulating the intermediate, staged integration of evidence. This neglect of the cognitive plausibility of *progressive integration* may lead to suboptimal performance and limited robustness, particularly for relations requiring multi-step reasoning or disambiguation against semantically similar alternatives.

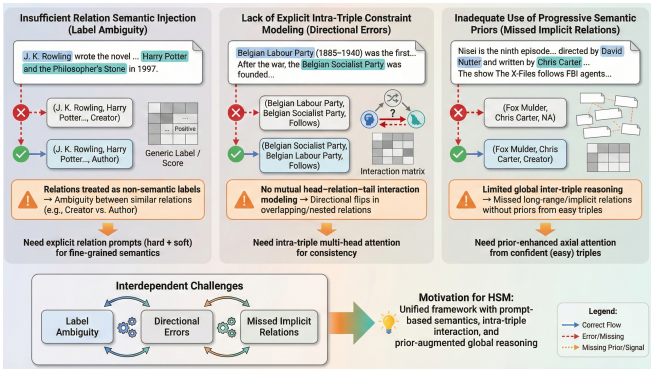

Figure 1: The challenge of error propagation in DocRE: directly predicting complex relations from noisy text can lead to cascading errors. (B) Our proposed staged cognitive model: reliable local semantic grounding (Stage 1) supports robust triple formation (Stage 2), which in turn provides stable priors for global reasoning (Stage 3), thereby mitigating error spread.

We hypothesize that explicitly structuring a computational model to mimic the **progressive coherence-building** observed in human comprehension will lead to more robust and accurate relational understanding. This hypothesis is grounded in cognitive theories of text processing (Kintsch, 1991) and problem-solving (Newell, Simon, et al., 1972), which emphasize *local-to-global integration* and *constraint satisfaction* as fundamental principles. A model that first establishes reliable local semantic bindings, then uses them to constrain the formation of higher-order relational structures, should be less prone to the error propagation that can occur when complex inferences are attempted without a solid foundation. To test this hypothesis, we propose **Holistic Semantic Modeling (HSM)**—a *hierarchical cognitive model* explicitly

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

designed to simulate progressive coherence building in relational understanding. HSM is structured as a cascade of three computational stages, each corresponding to a postulated level in the coherence-building process:

1. **Atomic Semantic Grounding**: The model enriches entity representations by fusing them with relational semantics via prompts, establishing a disambiguated, relation-aware "perceptual" layer.
2. **Local Coherence Formation**: It enforces internal consistency within candidate entity-relation-entity triples through attention-based binding, forming reliable local "chunks" or propositions.
3. **Global Coherence Integration**: It propagates constraints from these locally coherent chunks across the entire document via prior-enhanced axial attention, resolving ambiguities and inferring implicit relations to achieve global coherence.

This staged architecture embodies the "progressive" principle: each stage provides a scaffold for the next, and errors are contained and corrected at lower levels before they can corrupt higher-level reasoning. Our work makes two primary contributions:

- **Theoretical/Computational**: We formalize "progressive coherence building" as a hierarchical computational process for relational understanding and instantiate it in the HSM architecture. This provides a concrete, testable model of how local-to-global integration can be achieved in a neural system.
- **Empirical**: We demonstrate that this cognitively-inspired, staged model achieves state-of-the-art performance on three challenging DocRE benchmarks (DocRED, CDR, GDA). Crucially, ablation studies reveal a pattern of cumulative degradation—removing higher stages causes greater performance loss—which empirically validates the model's progressive dependency structure and its role in mitigating error propagation.

Related Work

Our work bridges advances in DocRE architectures and cognitive-inspired modeling strategies.

Document-level RE Architectures. Early sequence-based models (Yao et al., 2019) processed text flatly. Graph-based methods (Wang et al., 2020) introduced explicit structure but often require heuristic graph construction. Transformer-based models (B. Xu et al., 2021) leverage self-attention for global context but treat relations atomically. Table-based models (N. Zhang et al., n.d.; Zhou et al., 2020) organize entity pairs into a 2D grid, enabling joint modeling. HSM adopts the table structure but *explicitly decomposes* the modeling task into sequential stages within this framework, moving beyond monolithic encoding.

Cognitive Principles in NLP. The idea of stepwise or incremental processing has roots in cognitive architectures like ACT-R (Anderson & Lebiere, 2014) and has inspired NLP models for tasks like question answering (Geva et al., 2020) and reasoning (Wei et al., 2022). In DocRE, the concept of using "easy" instances to guide "hard" ones aligns with curriculum learning (Bengio et al., 2009) and self-training. HSM operationalizes this by making the "easy-to-hard" information flow an *architectural feature* via its staged pipeline and prior-propagation mechanism, rather than just a training schedule.

Method

The HSM framework (Fig.2) is designed as a cascade of three computational modules, each corresponding to a stage in our proposed cognitive model. The flow is predominantly sequential to enforce the dependency of complex reasoning on simpler, established representations.

Stage 1: Atomic Semantic Grounding via Prompt Injection

Objective: To produce entity and relation representations that are semantically rich and disambiguated at the *local level*, providing a solid foundation for later stages.

Cognitive Analogy: This stage corresponds to the initial perception and lexical access phase, where words are mapped to concepts. Ambiguity at this stage can propagate, so we enrich the input with prior relational knowledge.

Modeling: We enhance the standard entity encoder with two complementary sources of relation semantics:

- **Hard Prompts (Explicit Semantics)**: Hard-prompts are manually selected tokens treated like regular input words. Each token corresponds to a specific DocRED relation. For example, we might use "born in" for `P19 (place_of_birth)` and "founder" for `P112 (founded_by)`. Formally, we obtain

$$R_{HP} = \{xr_1,\ xr_2,\ \ldots,\ xr_K\} \tag{1}$$

  where each xr_k is the chosen hard-prompt token for relation r_k.
- **Soft Prompts (Learned Semantics)**: Soft-prompts are learned, continuous embeddings derived from labeled examples and, when available, relation descriptions. For each relation r_i, let

$$EP_i = \{(s_l, o_l)\}_{l=1}^{L_i} \tag{2}$$

  be all entity pairs annotated with r_i, and let $\mathbf{d}_i$ be the [CLS] representation of its textual description (if provided) via the same encoder. After encoding each sentence containing (s_l, o_l), we pool its token embeddings:

$$\mathbf{h}_{i,l} = \text{AvgPool}\big(\text{Encoder}(s_l, o_l)\big) \tag{3}$$

  We first aggregate the example-driven features:

$$\mathbf{u}_i = \frac{1}{L_i} \sum_{l=1}^{L_i} \mathbf{h}_{i,l} \tag{4}$$

  and then fuse with the description embedding:

$$\mathbf{hr}_i = \alpha\, \mathbf{u}_i + \beta\, \mathbf{d}_i \tag{5}$$

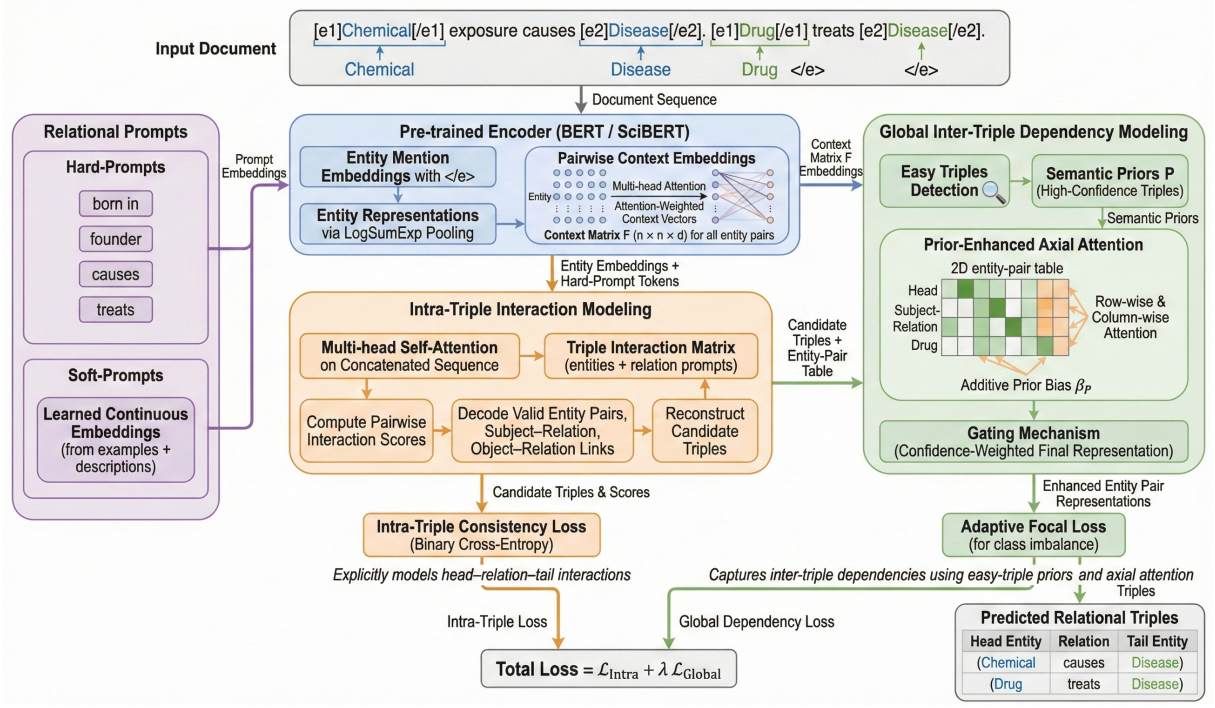

Figure 2: Our Holistic Semantic Model (HSM) framework, a *hierarchical cognitive model* explicitly designed to simulate progressive coherence building in relational understanding. Information flows sequentially, with each stage building upon the output of the previous one to minimize error propagation.

where α, β are learned fusion weights. The set of soft-prompts is

$$R_{SP} = \{\mathbf{hr}_1,\ \mathbf{hr}_2,\ \ldots,\ \mathbf{hr}_K\} \tag{6}$$

By combining discrete hard-prompts (e.g. "born in," "founder") with feature- and description-driven soft-prompts, our model enriches entity and relation representations with both explicit human cues and contextual semantic features. A pre-trained language model encoder (e.g., BERT(Devlin et al., 2019a)) processes the document augmented with these prompts, yielding context-aware representations $\mathbf{H}$ for all tokens. Entity representations $\mathbf{e}_i$ are obtained via LogSumExp pooling over their mentions. Crucially, these representations now carry relation-informed semantics *before* any triple-level reasoning begins.

Stage 2: Local Triple Formation via Intra-Triple Interaction

Objective: To bind the grounded entities and relations from Stage 1 into internally consistent candidate triples.

Cognitive Analogy: This mirrors the process of forming a basic proposition or "chunk" in working memory. The focus is on local compatibility: do the head, relation, and tail fit together?

Modeling:

Intra-triple Encoding We first concatenate each relation's hard-prompt with the input entity and encode the sequence (including both entity markers and prompt tokens) using a pretrained Transformer (e.g., BERT). Let $\mathbf{H} \in \mathbb{R}^{d\times L}$ be the output embeddings of the final layer. We project $\mathbf{H}$ to obtain queries, keys, and values:

$$\mathbf{Q} = \mathbf{W}_q\,\mathbf{H} + \mathbf{b}_q, \tag{7}$$

$$\mathbf{K} = \mathbf{W}_k\,\mathbf{H} + \mathbf{b}_k, \tag{8}$$

$$\mathbf{V} = \mathbf{W}_v\,\mathbf{H} + \mathbf{b}_v, \tag{9}$$

and compute multi-head self-attention:

$$\mathbf{A} = \text{softmax}\Big(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\Big)\,\mathbf{V}. \tag{10}$$

The resulting $\mathbf{A}$ captures pairwise interactions among all tokens (entities and prompts). We then integrate learned soft-prompts by fusing their embeddings with $\mathbf{A}$ to obtain a unified representation for each candidate triple.

Triple Interaction Matrix From the last Transformer layer's multi-head outputs $\{\mathbf{Q}_h, \mathbf{K}_h\}_{h=1}^{H}$, we compute the interaction score between any two tokens (entities or relations) as

$$\mathbf{I} = \sigma\Big(\frac{1}{H}\sum_{h=1}^{H}\frac{\mathbf{Q}_h\mathbf{K}_h^\top}{\sqrt{d}}\Big), \tag{11}$$

where $\sigma(\cdot)$ is the sigmoid function. Thresholding $\mathbf{I}_{i,j}$ yields a binary matrix indicating valid entity–entity or entity–relation links.

Intra-triplet Decoding We extract valid entity pairs from the "entity–entity" block of $\mathbf{I}$, and valid subject–relation and relation–object pairs from its off-diagonal blocks. By aligning these three sets—entity–entity, subject–relation, and relation–object—we reconstruct all plausible triples.

Module Training We supervise the interaction matrix $\mathbf{I}$ with a binary cross-entropy loss as intra-triple consistency loss over all pairs:

$$\mathcal{L}_{\text{IntraRE}} = -\frac{1}{(N+K)^2} \sum_{i=1}^{N+K} \sum_{j=1}^{N+K} \Big[y_{i,j} \log I_{i,j} + (1 - y_{i,j}) \log(1 - I_{i,j}) \Big]. \tag{12}$$

where $y_{i,j} \in \{0, 1\}$ is the ground-truth label for position (i, j).

Stage 3: Global Structure Refinement via Prior-Enhanced Reasoning

Objective: To resolve remaining ambiguities and infer implicit relations by leveraging the reliable local triples from Stage 2 as constraints for global document-level reasoning.

Cognitive Analogy: This corresponds to integrative reasoning or "sense-making" across multiple propositions, using known facts to constrain inferences about unknown ones.

Modeling: Building on intra-triple semantic modeling, we construct candidate triples from predicted head–tail entities to capture plausible relations and enable global dependency modeling. Although prior work highlights the benefits of inter-triple correlation modeling, it often overlooks the rich semantic priors within individual triples. To address this, we propose a **semantic prior-enhanced axial attention** module that effectively captures global dependencies among relational triples by leveraging these priors.

Semantic Prior-Enhanced Axial Attention Rather than relying solely on the head and tail embeddings for relation classification, we propose a *semantic prior-enhanced two-hop axial attention* mechanism inspired by the notion of *Easy triples*—triples confidently predicted by the intra-triple module.

We define *semantic priors* based on high-confidence predictions, i.e., triples where the conditional probability exceeds a threshold α:

$$P(r \mid e_s, e_o) > \alpha \tag{13}$$

These are used to construct a weight matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$, where:

$$P_{s,o} = \begin{cases} P(r|e_s, e_o) & \text{if } (e_s, r, e_o) \text{ is an Easy triple} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

Given an $n \times n$ entity table, for any pair (e_s, e_o), axial attention attends to neighboring positions along both the row and the column—i.e., (e_s, e_i) and (e_i, e_o)—thereby enabling two-hop compositional reasoning.

The attention computation is modified to integrate semantic priors:

Horizontal (row-wise) attention:

$$r_h^{(s,o)} = g^{(s,o)} + \sum_{p=1}^{n} \text{softmax}_p \left(\underbrace{q_{(s,o)}^\top k_{(s,p)}}_{\text{content-based}} + \beta P_{s,p} \right) v_{(s,p)} \tag{15}$$

Vertical (column-wise) attention:

$$r_w^{(s,o)} = r_h^{(s,o)} + \sum_{p=1}^{n} \text{softmax}_p \left(\underbrace{q_{(s,o)}^\top k_{(p,o)}}_{\text{content-based}} + \beta P_{p,o} \right) v_{(p,o)} \tag{16}$$

Here, $q_{(i,j)} = W_Q g^{(i,j)}$, $k_{(i,j)} = W_K g^{(i,j)}$, and $v_{(i,j)} = W_V g^{(i,j)}$ are linear projections of the entity pair representation $g^{(i,j)}$, with learnable matrices $W_Q, W_K, W_V \in \mathbb{R}^{d \times d}$. The scalar β is a learnable scaling factor that balances the influence of semantic priors.

To regulate the final output based on prior confidence, we introduce a gating mechanism:

$$g_{(s,o)} = \sigma \left(\mathbf{W}_g \left[g^{(s,o)} \parallel \max_r P(r \mid e_s, e_o) \right] \right) \tag{17}$$

$$\tilde{r}_w^{(s,o)} = g_{(s,o)} \odot r_w^{(s,o)} \tag{18}$$

where σ is the sigmoid activation, $\| \cdot \|$ denotes vector concatenation, and $\odot$ represents element-wise multiplication.

This stage is trained with an Adaptive Focal Loss(Zhou et al., 2021) $\mathcal{L}_{\text{GlobalRE}}$ to handle the severe class and positive-negative imbalance in relation prediction.

Joint Training as Coordinated Learning

The three stages are trained jointly end-to-end, with a total loss:

$$\mathcal{L}\text{total} = \mathcal{L}_{\text{IntraRE}} + \lambda \mathcal{L}_{\text{GlobalRE}}. \tag{19}$$

While information flows sequentially during *inference* (Stage 1 → Stage 2 → Stage 3), *training* is joint. This allows gradient signals from the global objective to fine-tune the earlier stages, encouraging them to produce representations that are optimal for the final integrated task—a process akin to "analysis by synthesis" in cognitive models.

Experiment

Our experiments are designed not only to demonstrate overall performance but also to validate the core hypothesis: that a staged, progressive modeling approach mitigates error propagation and leads to more robust reasoning.

Datasets and Metrics

We use three standard DocRE benchmarks: DocRED (Yao et al., 2019) (general domain), CDR and GDA (Li et al.,

2016; Piñero et al., 2016) (biomedical domain).**DocRED** is a large-scale dataset constructed from Wikipedia, containing 5,053 documents and 97 relation types. It features an average of 12.6 relational facts per document, significantly more than typical sentence-level RE datasets.

CDR and **GDA** are biomedical domain datasets focusing on single-relation extraction: CDR targets Chemical-Induced-Disease relations between chemical and disease entities, while GDA covers Gene-Induced-Disease relations between gene and disease entities. The dataset statistics are shown in Table 1. We evaluated our model with Ign F1, and F1 following(Yao et al., 2019), Ign F1 measures the F1 score that excludes the relationship shared between the training set and the dev/test set.

Table 1: Statistics of the experimental datasets.

Statistics / Datasets	DocRED	CDR	GDA
# Train	3053	500	23353
# Dev	1000	500	5839
# Test	1000	500	1000
# Relation types	97	2	2
# Avg.# mentions per Ent	1.4	2.7	3.3
# Avg.# entities per Doc	19.5	7.6	5.4
# Avg.# sentences per Doc	8.0	9.7	10.2

Implementation Details

We implemented HSM with the Pytorch version of Huggingface Transformers[1]. We use cased BERT(Devlin et al., 2019b) as the encoder on DocRED(Yao et al., 2019), and SciBERT(Beltagy et al., 2019) on CDR(Li et al., 2016) and GDA(Piñero et al., 2016). The training process is based on Apex library. Our model is optimized with AdamW(Loshchilov & Hutter, 2017) using learning rates 2e-5, with a linear warmup for the first 6% steps followed by a linear decay to 0. We perform early stopping based on the F1 score on the development set. We set the entity-level relation matrix size $N = 48$, $\lambda = 0.1$, $\gamma = \alpha = 0.5$.

Main Results

On the benchmark dataset DocRED, we compared HSM with twelve baseline models. The experiment results are shown in Table 2. We compare the HSM model with seventeen baselines: Sequence based models include CNN(Yao et al., 2019), BiLSTM(Yao et al., 2019). GNN based models include GLRE(Wang et al., 2020) and HeterGSAN(W. Xu et al., 2021). Transformer based models include Coref-BERT(Ye et al., 2020), SSAN(B. Xu et al., 2021). Table based models include DocuNet(N. Zhang et al., 2021), EICS(S. Liu et al., 2023), REI(W. Liu, Zeng, et al., 2025), SEREPI(W. Liu, Xiao, et al., 2025).

[1]https://huggingface.co/docs/transformers/index

Table 2: Experimental results (%) on dev and test sets of the DocRED dataset. The best scores are in **bold**.

Model	Dev		Test	
	Ign F1	F1	Ign F1	F1
Sequence-based				
CNN	41.58	43.45	40.33	42.26
BiLSTM	48.87	50.94	48.78	51.06
Graph-based				
GLRE-BERT$_{base}$	-	-	55.40	57.40
HeterGSAN-BERT$_{base}$	58.13	60.18	57.12	59.45
Transformer-based				
Coref-BERT$_{base}$	55.32	57.51	54.54	56.96
SSAN-BERT$_{base}$	57.03	59.19	55.84	58.16
ATLOP-BERT$_{base}$	59.22	61.09	59.31	61.30
Table-based				
DocuNet-BERT$_{base}$	59.86	61.83	59.93	61.86
KDDocRE-BERT$_{base}$	60.08	62.03	60.04	62.08
EICS-BERT$_{base}$	59.83	61.92	59.53	61.87
REI-BERT$_{base}$	61.12	63.17	60.51	62.40
SEREPI-BERT$_{base}$	60.53	62.42	60.40	62.82
Our Model				
HSM-BERT$_{base}$	**61.50**	**63.70**	**61.80**	**63.90**

As shown in Table 2, HSM-BERT$_{base}$ achieves state-of-the-art results on the DocRED dataset, with **61.50%/63.70%** Ign F1/F1 on the dev set and **61.80%/63.90%** on the test set, surpassing the strongest baseline REI-BERT$_{base}$. These consistent gains demonstrate HSM's ability to model holistic semantics and fine-grained entity interactions, advancing the state of document-level relation extraction. HSM achieves SOTA, demonstrating the effectiveness of staged modeling on a large, complex dataset.

Table 3 displays the experimental results of HSM on the biomedical dataset, as well as the comparison with the twelve baseline model: CNN(Verga et al., 2018) and BRAN(Nguyen & Verspoor, 2018) are both sequence-based models. DHG(Z. Zhang et al., 2020) and GLRE(Wang et al., 2020) are all graph-based models. SSAN (B. Xu et al., 2021) are both transformer-based models. DocuNet(N. Zhang et al., 2021), DenseCCNet(L. Zhang & Cheng, 2022), REI(W. Liu, Zeng, et al., 2025) and SEREPI(W. Liu, Xiao, et al., 2025) are table-based models. HSM achieves state-of-the-art results with 78.8% F1 on CDR and 87.1% on GDA, outperforming the strongest baseline by +1.8% and +0.7% respectively. Notably, it shows substantial recall gains (+7.9% on CDR, +4.4% on GDA), demonstrating superior capability in capturing comprehensive relational evidence for biomedical relation extraction.

Ablation Study

The ablation study (Table 4) provides crucial evidence for our cognitive modeling thesis. The performance drop is **largest when the final, global reasoning stage (Stage 3) is removed** (-1.90 F1). This confirms that complex reasoning

Table 3: Test Precision (P), Recall (R) and F1 (%) on the CDR and GDA datasets. The best scores are shown in **bold**.

Model	CDR			GDA		
	P	R	F1	P	R	F1
Sequence-based						
BRAN	55.6	70.8	62.1	–	–	–
CNN	57.0	68.6	62.3	–	–	–
Graph-based						
DHG	61.8	70.5	65.9	80.8	85.5	83.1
GLRE	65.1	72.2	68.5	–	–	–
Transformer-based						
SSAN	–	–	68.7	–	–	83.7
ATLOP	–	–	69.4	–	–	83.9
Table-based						
REI	–	–	73.2	–	–	85.8
SEREPI	–	–	73.3	–	–	84.8
DocuNet	–	–	76.3	–	–	85.3
DenseCCNet	–	–	77.0	–	–	86.4
Our Model						
HSM	**78.2**	**79.5**	**78.8**	**84.8**	**89.6**	**87.1**

without the constraint propagation mechanism is highly error-prone. Removing the local structure formation signal (Intra-Triple Loss) also causes a significant drop (-1.15 F1). Removing components of the initial semantic grounding stage (prompts) causes smaller but consistent declines. This pattern of **increasing performance penalty for removing later, more complex stages** is exactly what the progressive coherence principle predicts: errors or omissions at the foundation propagate upward, causing greater harm at the highest level of integration.

Table 4: Ablation study on DocRED dev set. Performance degrades progressively as we remove foundational stages, confirming the staged design's role in error control.

Model Variant (Removed Component)	F1 ↓
Full HSM Model	63.70
- Stage 3: w/o Global Dependency Module	61.80 (-1.90)
- Stage 2 & 3 Implicit: w/o Intra-Triple Loss	62.55 (-1.15)
- Part of Stage 1: w/o Hard-Prompts	62.90 (-0.80)
- Part of Stage 1: w/o Desc.-based Soft-Prompts	63.10 (-0.60)
- Part of Stage 1: w/o Ex.-based Soft-Prompts	63.30 (-0.40)

Case Study

To illustrate how our model's three-stage cognitive architecture enables robust relational understanding, we analyze three representative examples from the DocRED development set. Each case highlights how a specific stage in our progressive coherence-building process corrects an error that occurs in a strong baseline model SEREPI (W. Liu, Xiao, et al., 2025).

Synthesis of Cognitive Insights: Collectively, these cases demonstrate how HSM's staged architecture controls error propagation—a core motivation of our cognitive modeling approach.

- In **Example 1**, a potential error is **prevented at the source** (Stage 1) through rich semantic grounding, avoiding the need for later correction.
- In **Example 2**, an error arising from structural ambiguity is **caught and corrected at an intermediate stage** (Stage 2) via internal consistency checks, preventing a directionally flawed triple from influencing global reasoning.
- In **Example 3**, the most complex error (missing a relation due to dispersed evidence) is **resolved at the highest stage** (Stage 3) by leveraging the reliable outputs of earlier stages as constraints for cross-context inference.

This cascading, stage-specific error mitigation mirrors the robustness of human comprehension, where misunderstandings are often resolved locally before they distort the global interpretation of a text. The failure of the strong baseline across all three cases underscores the limitation of models that do not explicitly implement such a progressive, hierarchical coherence-building mechanism.

Table 5: Case study of progressive coherence building in HSM. Each example demonstrates how a specific cognitive stage (Grounding, Formation, or Integration) resolves a distinct type of ambiguity, preventing error propagation. ✓ indicates correct prediction; × indicates error.

Document Snippet & Core Challenge	**Model / Stage Analyzed**	**Prediction & HSM's Cognitive Process**	**Cognitive Interpretation: Stage-Specific Resolution**
Example 1: *Challenge*: Distinguishing closely related relation types (*Author* vs. *Creator*). *Text*: "[0]J. K. Rowling wrote the novel Harry Potter and the Philosopher's Stone in 1997..."	SEREPI (Baseline)	(Rowling, HP Stone, Creator) ×	Flat classification fails due to insufficient semantic grounding. The model conflates similar relational concepts without access to relation-specific semantics at the input level.
	HSM Stage 1: Grounding	Prompt injection enriches Rowling with semantics of *person*, *writer*, and cues for `P50 (Author)`.	**Stage 1 Success**: Rich semantic grounding prevents an early misclassification.
	HSM Final Output	(Rowling, HP Stone, Author) ✓	
Example 2: Directional & Internal Consistency *Challenge*: Inferring correct direction of relation (*Follows*). *Text*: "[0]Belgian Labour Party (1885–1940)... [2] After the war, the Belgian Socialist Party was founded in 1945..."	SEREPI (Baseline)	(BLP, BSP, Follows) ×	Without internal consistency checks, the model produces directionally inverted triples that violate temporal constraints.
	HSM Stage 2: Formation	Intra-triple attention detects incompatibility: older BLP cannot *follow* newer BSP.	
	HSM Final Output	(BSP, BLP, Follows) ✓	**Stage 2 Success**: Local consistency checking corrects structural errors.
Example 3: Cross-Context Evidence Integration *Challenge*: Inferring a relation from dispersed evidence. *Text*: "...[1]Nisei is an episode of The X-Files. ... [3] It was written by Chris Carter. ... [8] The show follows Fox Mulder..."	SEREPI (Baseline)	(Mulder, Carter, NA) ×	Isolated entity-pair analysis fails to aggregate multi-hop evidence distributed across contexts.
	HSM Stages 1 & 2	Establishes reliable local triples involving The X-Files as an anchor entity.	
	HSM Stage 3: Integration	Global propagation infers (Mulder, Carter, `P170 (Creator)`) ✓	**Stage 3 Success**: Global integration aggregates dispersed evidence via relational analogy.

Conclusion

We have presented the Holistic Semantic Modeling (HSM) framework for document-level relation extraction, explicitly designed around the cognitive principle of (stepwise thinking, from simple to difficult). By decomposing the complex DocRE task into three stages—Atomic Semantic Grounding, Local Triple Formation, and Global Structure Refinement—HSM builds a robust reasoning process where each stage provides a stable foundation for the next, effectively mitigating error propagation. Our experimental results demonstrate state-of-the-art performance, and our ablation studies confirm the cumulative, staged nature of the model's contributions.

Acknowledgments

This work was supported by the National Key Research and Development Program of China under Grant No.2023YFC3304404.

References

Anderson, J. R., & Lebiere, C. J. (2014). *The atomic components of thought*. Psychology Press.

Beltagy, I., Lo, K., & Cohan, A. (2019, November). SciBERT: A pretrained language model for scientific text. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Emnlp-ijcnlp* (pp. 3615–3620). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1371

Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. *Proceedings of the 26th Annual International Conference on Machine Learning*, 41–48. https://doi.org/10.1145/1553374.1553380

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019a). Bert: Pre-training of deep bidirectional transformers for language understanding. *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)*, 4171–4186.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019b, June). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Naacl* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Geva, M., Gupta, A., & Berant, J. (2020). Injecting numerical reasoning skills into language models. *arXiv preprint arXiv:2004.04487*.

Kintsch, W. (1991). The role of knowledge in discourse comprehension: A construction-integration model. In G. Stelmach & P. Vroon (Eds.), *Text and text processing* (pp. 107–153, Vol. 79). North-Holland. https://doi.org/https://doi.org/10.1016/S0166-4115(08)61551-4

Li, J., Sun, Y., Johnson, R. J., Sciaky, D., Wei, C.-H., Leaman, R., Davis, A. P., Mattingly, C. J., Wiegers, T. C., & Lu, Z. (2016). Biocreative v cdr task corpus: A resource for chemical disease relation extraction. *Database*, *2016*.

Liu, S., Shen, X., Liu, T., & Lan, M. (2023). Document-level relation extraction with entity interaction and commonsense knowledge. *2023 International Joint Conference on Neural Networks (IJCNN)*, 1–9.

Liu, W., Xiao, Y., Cheng, S., Zeng, D., Zhou, L., Kong, W., Zhang, M., & Chen, W. (2025). Document-level relation extraction with structural encoding and entity-pair-level information interaction. *Expert Systems with Applications*, *268*, 126099.

Liu, W., Zeng, D., Zhou, L., Xiao, Y., Zhang, M., & Chen, W. (2025). Enhancing document-level relation extraction through entity-pair-level interaction modeling. *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5.

Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Newell, A., Simon, H. A., et al. (1972). *Human problem solving* (Vol. 104). Prentice-hall Englewood Cliffs, NJ.

Nguyen, D. Q., & Verspoor, K. (2018). Convolutional neural networks for chemical-disease relation extraction are improved with character-based word embeddings. *arXiv preprint arXiv:1805.10586*.

Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., & Furlong, L. I. (2016). Disgenet: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research*, gkw943.

Verga, P., Strubell, E., & McCallum, A. (2018). Simultaneously self-attending to all mentions for full-abstract biological relation extraction. *arXiv preprint arXiv:1802.10569*.

Wang, D., Hu, W., Cao, E., & Sun, W. (2020). Global-to-local neural networks for document-level relation extraction. *arXiv preprint arXiv:2009.10359*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824–24837.

Xu, B., Wang, Q., Lyu, Y., Zhu, Y., & Mao, Z. (2021). Entity structure within and throughout: Modeling mention dependencies for document-level relation extraction. *Proceedings of the AAAI conference on artificial intelligence*, *35*(16), 14149–14157.

Xu, W., Chen, K., & Zhao, T. (2021). Document-level relation extraction with reconstruction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(16), 14167–14175.

Yao, Y., Ye, D., Li, P., Han, X., Lin, Y., Liu, Z., Liu, Z., Huang, L., Zhou, J., & Sun, M. (2019, July). DocRED: A large-scale document-level relation extraction dataset. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Acl* (pp. 764–777). ACL. https://doi.org/10.18653/v1/P19-1074

Ye, D., Lin, Y., Du, J., Liu, Z., Li, P., Sun, M., & Liu, Z. (2020). Coreferential reasoning learning for language representation. *arXiv preprint arXiv:2004.06870*.

Zhang, L., & Cheng, Y. (2022). A densely connected criss-cross attention network for document-level relation extraction. *arXiv preprint arXiv:2203.13953*.

Zhang, N., Chen, X., Xie, X., Deng, S., Tan, C., Chen, M., Huang, F., Si, L., & Chen, H. (2021). Document-level relation extraction as semantic segmentation. *arXiv preprint arXiv:2106.03618*.

Zhang, N., Chen, X., Xie, X., Deng, S., Tan, C., Chen, M., Huang, F., Si, L., Chen, H., & Center, H. I. (n.d.). Document-level relation extraction as semantic segmentation.

Zhang, Z., Yu, B., Shu, X., Liu, T., Tang, H., Yubin, W., & Guo, L. (2020). Document-level relation extraction with dual-tier heterogeneous graph. *Proceedings of the 28th International Conference on Computational Linguistics*, 1630–1641.

Zhou, W., Huang, K., Ma, T., & Huang, J. (2020). Document-level relation extraction with adaptive thresholding and localized context pooling. https://arxiv.org/abs/2010.11304

Zhou, W., Huang, K., Ma, T., & Huang, J. (2021). Document-level relation extraction with adaptive thresholding and localized context pooling. *Proceedings of the AAAI conference on artificial intelligence*, *35*(16), 14612–14620.