Modal Dependency Parsing as Structured Prediction over Source–Cue Scopes

Jayeol Chun
Brandeis University
415 South Street, Waltham, MA 02453
jchun@brandeis.edu

Nianwen Xue
Brandeis University
415 South Street, Waltham, MA 02453
xuen@brandeis.edu

Abstract

Modal dependency parsing—the task of identifying a semantic graph that represents who is responsible for an event-centered claim and with what degree of certainty—relies on recognizing source-introducing cues and correctly linking them to their associated content. However, prior work has largely focused on identifying sources only, treating cue expressions and their modal coverage as auxiliary signals. In this work, we propose a structured prediction framework that leverages large language models (LLMs) to explicitly identify source–cue pairs as well as their respective scope, which together define the modal contexts governing downstream source attribution for events. By concentrating learning at the source-cue level and constraining event-level decisions to a small, scope-defined candidate set, our top-down approach enables inference over a reduced candidate space in long, event-rich documents. Experiments show this approach surpasses prior state-of-the-art results by 3 and 4% for English and Chinese datasets, respectively.

1 Introduction

Modal Dependency Parsing (MDP) aims to identify and represent expressions of modality—such as belief, obligation, intention, and uncertainty—and to determine how these modal meanings scope over events in text. Rather than treating modality as a flat label, MDP represents modal relations as structured dependencies that together form a document-level semantic dependency structure, known as Modal Dependency Graph (MDG) (Vigus et al., 2019), capturing who holds a modal attitude over which events or states it governs. This structured view is designed to enable fine-grained semantic interpretation, supporting downstream tasks such as factuality assessment and discourse understanding (Yao et al., 2021).

Structurally, an MDG is a single-headed, directed tree whose nodes are either events or sources.

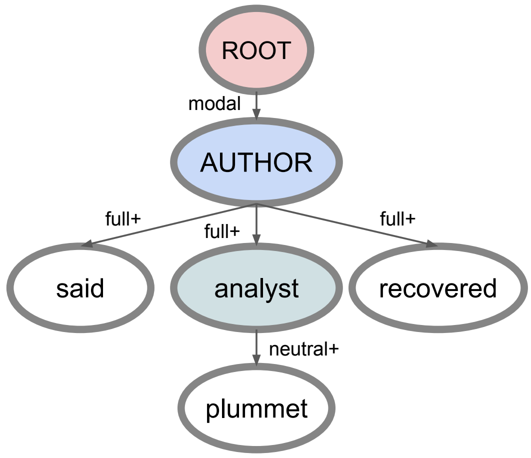


Figure 1: Example Modal Dependency Graph. '+' is a short-hand for 'affirmative' polarity value.

A source, or *conceiver*, is the entity responsible for introducing or endorsing an event-centered proposition, serving as the anchor of modal interpretation for the associated context. Edge direction encodes source attribution by linking a claim to the entity responsible for its assertion, while the edge label shows epistemic strength (*full*, *partial* or *neutral*) and polarity values (*affirmative* or *negative*), motivated by FactBank annotation scheme (Saurí and Pustejovsky, 2009; Vigus et al., 2019).

Figure 1 shows a sample MDG for a document of two sentences:

(1) a. The analyst said the company stock might plummet.

b. It has since recovered.

Each MDG is headed by an abstract ROOT node, depicted as a light-red node at the top. The light-blue AUTHOR node corresponds to the principal writer of the document and acts as the default source for many events and statements in the text.

Indeed, the sub-graph rooted at the AUTHOR node captures the author's attributional commitments in the sample document above. Within this

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 29530–29540

space, the analyst's reported speech is instantiated as a "said" node, reflecting the author's full-affirmative commitment.

Through this reporting structure, the analyst is positioned as the source of the subsequently quoted content. Accordingly, when a speculative claim about a potential decline in the company's stock value is encountered, the attribution is assigned not to the author, but to the analyst. This is structurally encoded as a neutral ('*might*') modal label from the "analyst" node to the "plummet" event. MDG enforces global connectedness by requiring a conceiver node, such as the "analyst," to connect to a distinct source node responsible for its introduction (AUTHOR in this case)[1]. Finally, the claim that the stock value has recovered is attributed to the author, though the boundaries of the analyst's modal scope of "saying" predicate does admit some ambiguity.

The task of MDP therefore consists of the following sub-tasks:

1. Event Detection
2. Conceiver (Source) Detection
3. Edge Attachment
4. Edge Labeling

While the first two tasks concern node identification, source identification poses a greater challenge because events are largely context-independent, whereas source attribution requires resolving highly contextual information. The third and fourth tasks then construct a connected and labeled graph by establishing modal relations among the identified nodes.

In this work, we propose an LLM-based structured prediction framework that addresses the task's inherent multi-tasking complexity by front-loading the identification of source–cue pairs and their associated modal scopes, which in turn simplifies the subsequent edge prediction into a constrained—and often trivial—ranking setup. The key strengths of our approach are as follows:

1. Scoping naturally aligns with human processing of modal attribution.
2. Concentrating learning on source–cue–scope extraction yields a more tracktable top-down formulation that globally identifies source-introducing expressions, in contrast to bottom-up approaches that locally predict the best head for each event.
3. A single cross-lingual model can be trained to process both English and Chinese.

[1]The modal label between conceiver nodes is always full-affirmative.

To facilitate efficient fine-tuning, we augment the original MDG corpus for English with silver-generated cue and scope spans for both languages, which are not provided in the initial annotations. We additionally improve overall annotation quality through targeted quality control, such as the removal of media-related artifacts (e.g., headlines, image captions, and photo credits) that are often unrelated to the document's topic. Together, these enhancements yield a revised and augmented version of the MDG corpus that can be released alongside the trained model. Our experiments show that this proposed approach yields improvements of 3% and 4% on English and Chinese MDP, respectively.

2 Related Work

Research on factuality and modality in natural language has evolved from local, event-centered classification toward increasingly structured representations that explicitly encode source attribution and degrees of certainty. Early work on event factuality prediction (EFP) treated the task primarily as a labeling problem, assigning polarity or certainty values to individual predicates. These approaches ranged from rule-based systems exploiting lexical and semantic cues (Nairn et al., 2006; Lotan et al., 2013) to statistical (Diab et al., 2009; Saurí and Pustejovsky, 2012; Lee et al., 2015; Stanovsky et al., 2017) and neural (Rudinger et al., 2018; Qian et al., 2018; Pouran Ben Veyseh et al., 2019) models trained on annotated corpora. Despite these advances, most EFP models continued to operate at the level of isolated events, without explicitly modeling the relational structure linking events to their information sources.

Modal Dependency Structure (MDS) was introduced to address this limitation by representing modality and factuality as a dependency structure over events and conceivers (Vigus et al., 2019). Building on this representation, Yao et al. (2021) reformulated factuality assessment as modal dependency parsing and released the first large-scale English dataset annotated with modal dependency graphs. This shift reframed factuality as a structured prediction problem, in which events and

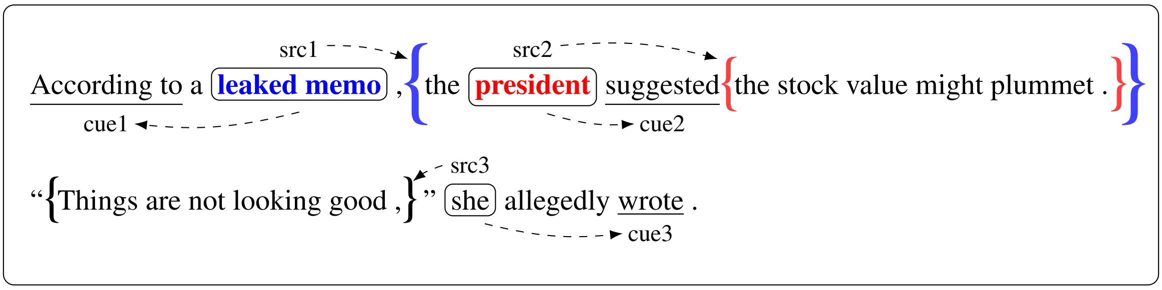


Figure 2: Source–Cue–Scope view of attributional structure. Scoped context is indicated by curly brackets.

sources are jointly embedded in a document-level graph, enabling more explicit reasoning about reported speech, attribution, and uncertainty.

Subsequent work extended this formulation to Chinese, producing an annotated dataset and reporting the first cross-lingual results (Liu and Xue, 2023; Yao et al., 2022). In parallel, modal dependency structures were incorporated into the Uniform Meaning Representation (UMR) framework (Van Gysel et al., 2021), where modality serves as a core document-level semantic layer alongside temporal relations and coreference. As a result, MDP has become an increasingly important subtask whose accuracy directly affects downstream parsing performance (Chun and Xue, 2024).

To date, the strongest reported results on MDP have been achieved by treating the task as a variant of sparse biaffine dependency parsing over tokens, which enables an efficient global inference over an entire document (Chun and Xue, 2025). A key strength of the biaffine parsers lies in their ability to perform document-level inference in a single forward pass, while maintaining strong inductive biases for head selection and relation labeling. By scoring head–dependent pairs directly, biaffine parsers can align local attachment decisions with global structural requirements like connectivity.

However, because conceiver node spans are not pre-identified but rather emerge jointly along with edge generation via a specially interpreted “SPAN” relation, edge-generation must encode two fundamentally different aspects of the task: (1) conceiver span boundary detection and (2) relational attachment. This joint representation complicates conceiver span recovery in a conventional dependency-based bottom-up formulations, as conceivers are highly context-dependent, often implicit, and difficult to localize. In contrast, our approach pre-identifies conceiver-related structure and decouples it from edge prediction, enabling a top-down formulation that prioritizes the discovery of source-introducing expressions and their scopes. The code is available at `https://github.com/umr4nlp/umrlib`.

3 Approach

3.1 Motivation

Modal expressions—such as quotations, reports, beliefs and intentions—are ubiquitous in natural language discourse (Palmer, 2001; Aikhenvald, 2004; Prasad et al., 2008; Semino and Short, 2011). Linguistically, these constructions invite a scoped interpretation in which readers first identify a source and the cue that introduces modal content, allowing for the incremental formation of distinct, nested contextual scopes. A defining characteristic of such constructions is that the embedded modal content is anchored to the nearest compatible source, yielding a preference for locally anchored interpretations in cases of nesting or ambiguity (Asher and Lascarides, 2003; Portner, 2009).

The first sentence in Figure 2 illustrates a nested attributional configuration, presenting source–cue–scope view of a sample document:

(2) a. According to a leaked memo, the president suggested the stock value might plummet.

b. “Things are not looking good,” she allegedly wrote.

Focusing on the event “plummet”, the outermost scope of the “leaked memo” and the inner scope of the “president” form a containment hierarchy, wherein the inner scope inherits the contextual constraints of the enclosing scope while providing more specific attributional commitment. In such configurations, anchoring to the innermost scope (the “president”) ensures that attribution is assigned to the *most specific* conceiver who directly and most immediately commits to the embedded

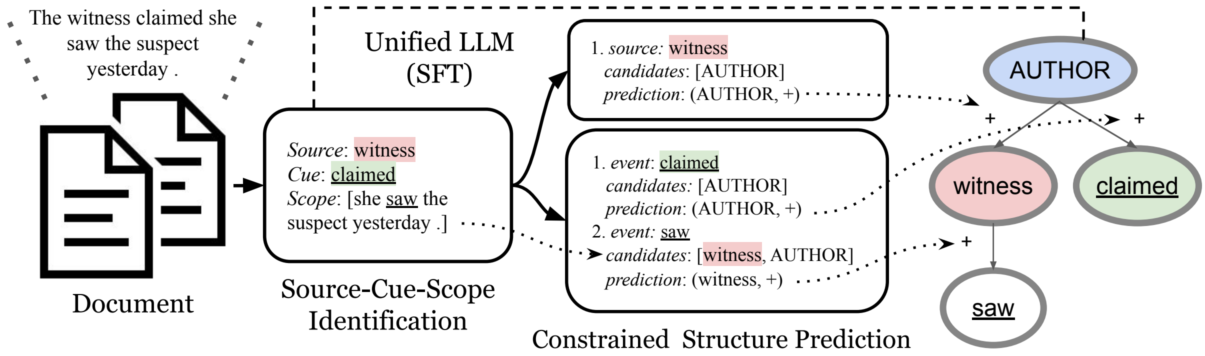


Figure 3: Pipeline of our proposed approach. Predictions are generated sequentially but trained jointly. + edge label is an abbreviation for "full-affirmative."

content, which naturally aligns with the single-head constraint of modal dependency parsing.

Finally, the event "wrote" in the second sentence is not covered by any of the currently identified source–cue scopes. In such cases, attribution defaults to AUTHOR; however, in passive or impersonal constructions, attachment may instead target a NULL abstract node[2]. It is worth noting this decision is not determined by surface proximity to nearby cues, but by whether the event falls within the coverage of pre-established attributional scopes.

This example highlights the central motivation of our approach—once attributional scopes are correctly identified, edge attachment reduces to a straightforward, scope-constrained decision, where each event can attach to the most specific enclosing scope or default to generic abstract node when no such scope exists. It must be noted that sources themselves must be assigned an appropriate attributional parent. Importantly, sources themselves must also be assigned an attributional parent as well, since the introduction of a new attributional context must be licensed by some conceiving entity. Finally, each attachment is labeled with the appropriate modal relation.

3.2 Task Formulation

Let a document input be represented as a sequence of tokens:

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

We formulate MDP as a top-down structured prediction problem that decomposes attribution modeling into two stages: (1) identifying source–cue–scope structures, then (2) deriving event-level attribution under explicit structural constraints.

3.2.1 Attributional Structure

We model attribution via a set of source–cue–scope triples:

$$\mathcal{S} = \{(s, c, \sigma)\}$$

where:

- s is a **source span**, corresponding to a conceiver or information source.
- c is a **cue span**.
- $\sigma \subseteq \{1, \ldots, n\}$ is a **scope span**, delimiting the region of text governed by the attribution introduced by c.

Each scope σ induces a partial order over attributional contexts via set containment. For any two scopes σ_i and σ_j, we write $\sigma_i \prec \sigma_j$ if $\sigma_i \subset \sigma_j$, yielding a hierarchy of nested attributional scopes.

3.2.2 Event Identification and Attribution

Event mentions are represented as a set of spans:

$$\mathcal{E} = \{e_1, \ldots, e_m\}$$

which are predicted jointly with $\mathcal{S}$.

Given the predicted source–cue–scope triples $\mathcal{S}$ and the set of detected events $\mathcal{E}$, the final task is to attach each event $e \in \mathcal{E}$ and source $s \in \mathcal{S}$ to an appropriate attributional context governed by another source by selecting a outgoing modal edge from a constrained candidate set. Both cases are handled using the same scope-constrained attachment mechanism, with the abstract AUTHOR and NULL node implicitly considered as valid candidates in all settings, ensuring that the candidate set is never empty.

[2]NULL node structurally denotes the absence of an explicitly mentioned source.

Specifically, for each node $x \in \mathcal{E} \cup \mathcal{S}$, we define a candidate set:

$$\mathcal{T}(x) = \{\, t = (s, c, \sigma) \in \mathcal{T} \mid x \in \sigma \,\},$$

consisting of all source–cue–scope triples whose modal scope contains the node. Edge attachment is then formulated as a constrained ranking problem over $\mathcal{T}(x)$, where the model selects the most compatible source-cue context for each node.

Given predicted attributional structure $\mathcal{S}$ and event spans $\mathcal{E}$, event-level attribution is derived as follows. For each event $e \in \mathcal{E}$, define the set of covering scopes:

$$\mathcal{S}(e) = \{(s, c, \sigma) \in \mathcal{S} \mid e \subseteq \sigma\}.$$

If $\mathcal{S}(e)$ is non-empty, the event is anchored to the *most specific* source:

$$s^* = \arg \min_{(s,c,\sigma) \in \mathcal{S}(e)} |\sigma|,$$

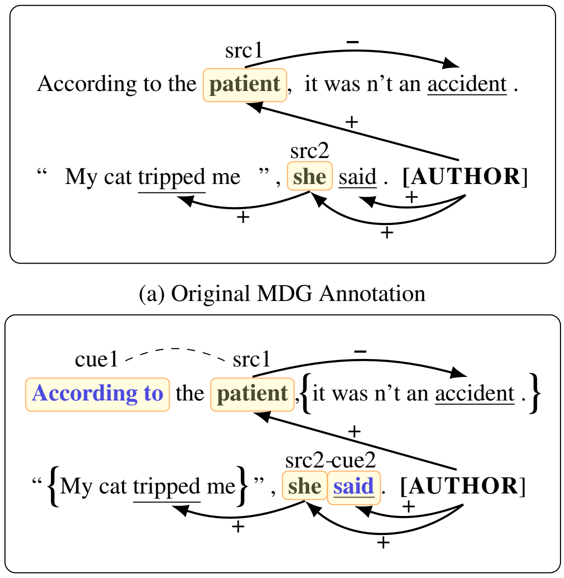
corresponding to the minimal enclosing attributional scope. If no attributional scope covers e, the event is anchored to AUTHOR, or to a distinguished NULL abstract node when no explicit source is present.

Under this formulation, the learning objective in a full supervised fine-tuning (SFT) setting is to jointly predict event spans $\mathcal{E}$ and attributional structure $\mathcal{S}$ from $\mathbf{x}$.

Although the target-level attribution admits a deterministic derivation by anchoring each target to the most specific enclosing scope or defaulting to AUTHOR/NULL, in practice we retain attribution as a learned prediction target, including both the modal edge attachment and its corresponding label. This design choice reflects the fact that annotation practices may deviate from strictly scope-based heuristics, for example in cases of implicit attribution or annotator preference. By learning target-level attribution edges and labels, including source-to-source relations, the model can capture such annotation behavior while remaining guided by the structural constraints imposed by scope containment.

3.3 Deriving Silver Cue-Scope Annotations

The original MDG corpora for English and Chinese do not explicitly annotate source-introducing cues or their associated modal scopes (Yao et al., 2021; Liu and Xue, 2023). To make this information available for efficient fine-tuning, we derive an auxiliary cue-scope annotation layer by querying a larger, off-the-shelf LLM reasoner. Leveraging the existing annotations that identify source spans, the model is prompted to detect source-introducing expressions, such as reporting predicates and prepositional constructions, and to delineate the contiguous span over which each cue introduces modal content. This process yields two alternative scope boundaries: ‘scope_min’, which adopts the tightest plausible boundary, and ‘scope_max’, which permits a broader interpretation of the attributional scope. This derivation is performed offline as a one-time preprocessing step.



(a) Original MDG Annotation

(b) Augmented Annotation with Source-Cue-Scope Triples

Figure 4: Original vs. Source-Cue-Scope Augmented MDG Annotation for a sample document of two sentences: ‘*According to the patient, it wasn’t an accident. "My cat tripped me", she said.*’ $+$ and $-$ denote “full-affirmative” and “full-negative” edge labels, respectively. AUTHOR is an abstract node that is introduced for visual clarity.

Figure 4 contrasts the original MDG annotation entry with the cue- and scope-augmented annotation for a two-sentence example. Whereas the original annotation leaves the source-introducing cues and the corresponding modal scope unmarked, the augmented annotation makes these elements explicit, thereby fully revealing the underlying attributional structure. These silver source–cue–scope annotations provide well-defined anchor points for loss computation during fine-tuning, enabling efficient supervision over MDP-specific structured prediction behavior.

English	Train	Dev	Test	Chinese	Train	Dev	Test
Documents	289	32	32	Documents	237	30	30
Sentences	6,825 (6,344)	740 (696)	759 (719)	Sentences	3,187	398	366
Tokens	151,487 (146,057)	17,308 (17,792)	17,177 (16,753)	Tokens	79,809	10,352	10,053
Conceivers	2,344 (2,310)	298 (296)	296 (293)	Conceivers	879	136	116
Events	19,541 (18,872)	2,307 (2,244)	2,168 (2,121)	Events	11,679	1,464	1,318
By Edge Label				*By Edge Label*			
`full +`	18,425 (17,838)	2,205 (2,153)	2,077 (2,036)	`full +`	10,879	1,383	1,257
`partial +`	1,292 (1,246)	165 (159)	158 (156)	`full -`	298	45	31
`neutral +`	1,368 (1,304)	136 (130)	140 (133)	`partial +`	919	103	101
`full -`	800 (785)	99 (98)	89	`neutral +`	429	64	45
`partial -`	-	-	-	`partial -`	26	0	0
`neutral -`	-	-	-	`neutral -`	7	0	0
Total	21,885 (21,173)	2,605 (2,540)	2,464 (2,414)	**Total**	12,558	1,595	1,434

Table 1: Summary statistics of English and Chinese modal dependency corpora. The abstract conceiver nodes AUTHOR and NULL, along with their "modal" edge to the ROOT, are not counted in the number of conceivers and edge-label, respectively. English corpus does not contain any annotations of `partial-negative` and `neutral-negative`.

3.4 Supervised Fine-Tuning Setup

We adopt a full supervised fine-tuning (SFT) regime to train the model to jointly predict attributional structure and target-level attribution. Fine-tuning is performed on the original MDG corpora, as well as the augmented corpus described above, leveraging silver annotations of cue–scope spans in addition to the original MDG annotations.

Input Representation. Each training instance consists of a pre-tokenized document input, accompanied by a token-level mapping that assigns a unique identifier to each token. This representation allows all span boundaries, including events, sources, cues, and scopes, to be expressed directly in a shared token index space, ensuring unambiguous alignment between model predictions and gold annotations.

Prediction Targets. Under SFT, the model is trained to primarily predict: (i) source spans, (ii) cue spans, (iii) scope spans, and (iv) attributional edges and labels over targets (events and sources). These fields are to be generated according to a fixed ordering as enumerated, so that it reflects the vertical factorization of modal dependency parsing—specifically, attributional structures at the source–cue–scope level are predicted first and subsequently used to constrain event- and source-level attribution decisions.

Training Objective. The training objective combines losses over span prediction and attributional edge prediction. Span-level losses supervise the identification of source, cue, and scope spans, while attributional losses supervise both the attachment decisions and their associated labels. Although attributional edges may be derived deterministically from scope containment in principle, they are retained as supervised targets to capture annotation conventions and variability present in the data. These components are jointly optimized as a single global loss over the training corpus.

Fine-Tuning Setup We fine-tune a reasoning-capable base LLM using a parameter-efficient supervised fine-tuning (SFT) regime based on QLoRA (Dettmers et al., 2023). This setup enables efficient adaptation by updating only a small subset of parameters while preserving the pre-trained model's representational capacity. Leveraging a base model with strong reasoning priors is particularly well suited to MDP, as the task requires integrating span-level decisions with hierarchical attributional structure under global constraints. Through SFT, these reasoning capabilities are redirected toward disciplined, structure-aware prediction behavior rather than free-form generation.

The fine-tuning prompt provides minimal, task-specific instructions that define the prediction order and output format. Specifically, the model is instructed to (i) identify source–cue pairs and their associated modal scopes using token indices, and (ii) generate attributional edges for target nodes under the constraint that each target attaches to the most specific enclosing scope or defaults to AUTHOR/NULL when no scope applies.

Multilingual Training. A single LLM is fine-tuned jointly on English and Chinese data using a shared set of prediction targets and supervision sig-

Models	Split	English			Chinese		
		Event	Conceiver	Parsing	Event	Conceiver	Parsing
Biaffine	Dev	93.0	73.1	74.5	87.2	**89.1**	68.6
	Test	91.8	74.7	73.3	87.5	87.3	66.7
Biaffine + Silver	Dev	93.3	72.9	74.5	87.5	88.8	69.0
	Test	91.5	74.2	73.5	87.9	87.4	67.3
Our Baseline	Dev	90.7	70.3	72.2	88.0	86.5	67.1
	Test	91.3	71.1	69.3	88.1	87.9	66.9
w/ Src-Cue-Scope	Dev	**93.3**	**74.3**	**77.7**	**88.4**	88.5	**72.1**
	Test	**91.8**	**76.3**	**76.2**	**89.9**	**88.0**	**71.1**

Table 2: Experimental results showing Event and Conceiver identification and Parsing micro-F score. The highest values are highlighted in bold. Empty values indicate unreported results. English results in the parenthesis denote evaluation on revised MDG corpus.

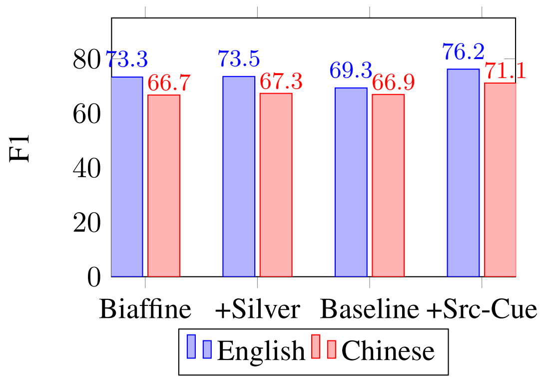


Figure 5: F1 Parsing performance across models on Test

nals. While our formulation is inherently language-agnostic, its effectiveness is necessarily bounded by the linguistic exposure and representational capacity of the underlying base LLM model, particularly given the relatively small size of the annotated corpora. To this end, we adopt a bilingual base from the Qwen family (Qwen et al., 2025).

4 Experiments

4.1 Corpus

The training, evaluation and test data for English are sourced from the modal dependency corpus (Yao et al., 2021), with annotations provided by crowd-sourced workers. For Chinese, the corresponding data are obtained from Liu and Xue (2023). Both corpora are drawn from the newswire domain, with the English dataset containing a substantially higher proportion of COVID-19-related coverage.

Our revised and augmented English MDG corpus preserves the original train/dev/test split. However, because the development and test sets have been modified, results on these splits are reported to establish new reference baselines for future comparison. Accordingly, results on the revised development and test sets are not intended for direct comparison with previously reported state-of-the-art numbers.

4.2 Model Configuration

In order to obtain silver annotations of cues and modal scopes, we query DeepSeek (DeepSeek-AI et al., 2025) via API requests.

Our choice of the base reasoning LLM to fine-tune is a 7B-distilled version of DeepSeek-R1 (DeepSeek-AI, 2025) based on Qwen 2.5 (Qwen et al., 2025).

4.3 Baselines

We compare our approach with a biaffine dependency parser (Chun and Xue, 2025), whose best results are obtained by using bi-directional Chinese-English translation for data augmentation. We further establish an in-house baseline that uses the same underlying LLM, but omits the structural factorization of source–cue–scopes, instead directly generating the entire MDG as JSON. This is labeled ‘Our Baseline.’

4.4 Results

Our experimental results are shown in Table 2 as micro F-score of single-best runs. Overall, the proposed approach with source–cue–scope modeling achieves the strongest performance across most settings. On the test set, our model improves parsing F1 from 73.5 to 76.2 for English and from 67.3 to 71.1 for Chinese over the Biaffine+Silver baseline, establishing new state-of-the-art results in both languages. Similar improvements are observed for Conceiver identification, where the model reaches

76.3 (English) and 88.0 (Chinese), outperforming all baselines.

We observe larger gains on the Chinese data, also reflected in Figure 5, suggesting that although our source–cue–scope modeling is inherently language-agnostic, it may offer additional benefits for Chinese compared to existing approaches. One plausible explanation is the reduced reliance on fragmented context encoding. The biaffine models for Chinese employ `xlm-roberta-base` encoder (Conneau et al., 2020) with the context window of 512 tokens, requiring sliding-window processing that can disrupt long-range dependencies critical for document-level parsing. In contrast, our approach leverages models with substantially larger context capacity, enabling more coherent global reasoning. The implementation details are provided in Appendix A.

5 Analysis

5.1 Against the Baselines

Our baseline of directly producing the full-fledged MDG as a single JSON object represents a strong but unconstrained generation setting, in which the model must jointly predict events, sources, attachments, and labels through sequence-level decoding. Despite using the same underlying LLM, this baseline consistently underperforms our structured approach as well as the biaffine parser. This gap highlights the difficulty of inducing graph-structured outputs through free-form serialization.

In particular, conceiver identification is notably weaker. In this setting, conceiver spans are predicted without explicit constraints over the token space, and the model must additionally infer how a source itself participates in a higher-order modal relation where it acts as a child. In practice, the baseline tends to default to generic anchors like AUTHOR, leading to lower conceiver recall and reduced downstream edge attachment accuracy.

5.2 Impact of Corpus Artifacts and Data Augmentation

We observe that the original MDG corpus contains preprocessing artifacts that complicate reliable training and evaluation. A prominent issue is the inclusion of media-related content—such as headlines from unrelated articles or captions of embedded media—which can introduce abrupt topic shifts that do not belong to the original document narrative. While this issue is largely absent in the Chinese data, it is more pronounced in the English portion of the corpus: 481 out of 6,820 training sentences (~7%) were flagged, manually verified, and subsequently removed in the revised dataset. Additional issues include token/phrase omissions and mis-tokenization,

	Event	Conceiver	Parsing
Dev	92.9	74.1	78.6
Test	91.4	76.5	76.9

Table 3: Results on the revised English MDG corpus

These artifacts naturally propagate into the token indexing process, yielding noisy mappings over which subsequent structured prediction must operate. As a result, attribution errors in the original corpus may not always be attributable to modeling limitations, but may instead reflect data inconsistencies that introduce noise into the training and evaluation process.

To this end, we report additional results in Table 3 conducted on the revised corpora, where the media-artifacts have been removed. Under this setting, we observe that event and conceiver identification performance remains largely unchanged, while the overall parsing F1 improves. This pattern suggests that corpus cleanup primarily benefits downstream attributional consistency and attachment quality, rather than altering the model's ability to detect individual nodes.

5.3 Sensitivity to Scope Boundary Choice

We analyze the effect of scope boundary selection by comparing two alternative configurations derived from the same source–cue annotations: 'scope_min' (default) and 'scope_max'. The 'scope_min' setting adopts the tightest plausible boundary around each source–cue pair, prioritizing precision in delimiting attributional context, while 'scope_max' allows broader scope extension to include potentially relevant downstream material.

Empirically, we observe that under 'scope_min', approximately **79%** of gold attributional targets are fully contained within the predicted scope as its inner context, compared to **75%** under 'scope_max'. This reduction reflects the tendency of broader scopes to introduce competing enclosing contexts, which weakens the alignment between innermost scope containment and gold attribution.

5.4 Cross-Lingual Comparison

A key distinction between English and Chinese attribution lies in the extent to which attribution is recovered from discourse context rather than signaled through overt linguistic cues. Chinese more readily permits dropped pronouns and implicit subjects, so that modal scope is often inferred from surrounding context rather than relying on surface cues.

Notably, the Chinese MDG dataset contains no instances of NULL connections, suggesting a consistent annotation practice where annotators resolve attribution to an implicit or previously introduced source, rather than marking it as unspecified. This contrasts with English, where passive constructions (e.g., "it was reported that…") are often annotated with NULL to reflect deliberate source suppression. In such settings, explicit scope modeling may help reveal the implicit discourse-level attribution, consistent with larger gains for Chinese.

6 Conclusion

In this work, we present a prediction framework that formulates modal dependency parsing as a top-down, structure-first problem. By explicitly modeling source–cue pairs and their associated modal scopes, our approach decouples attributional identification from event-level edge attachment, substantially simplifying downstream attachment decisions. In addition, we introduce an augmented version of the MDG corpora with explicit source–cue–scope annotations and targeted quality control for English, providing more reliable supervision and supporting the establishment of more robust baselines. Empirical results on English and Chinese MDP demonstrate consistent improvements of 3% and 4%, respectively.

Acknowledgment

This work is supported by grants from the CNS Division of National Science Foundation (Awards no: NSF_2213804) entitled "Building a Broad Infrastructure for Uniform Meaning Representations". Any opinions, findings, conclusions or recommendations expressed in this material do not necessarily reflect the views of NSF.

Limitations

Our model operates on pre-tokenized input, reflecting the fact that the MDG corpora for both English and Chinese provide tokenized text with token-aligned span annotations. In practical deployment on raw text, an additional tokenization and sentence-segmentation pipeline (e.g., CoreNLP (Manning et al., 2014) or an equivalent tool) is therefore required to produce compatible inputs. Because tokenization errors may propagate to downstream, model performance on real-world data is expected to be lower than the results reported in this work.

We assume that the modal scope associated with a source-cue pair covers a continuous span. While discontinuous scope can arise linguistically—for example, in interrupted quotations such as "The plan," he said, "will fail"—such cases can be safely absorbed within our framework.

Because the annotated English and Chinese data are limited to the newswire domain, our evaluation does not extend to other domains or languages. This constraint reflects the scope of the available annotated data, and we leave evaluation beyond English and Chinese newswire to future work.

References

Alexandra Y Aikhenvald. 2004. *Evidentiality*. Oxford University Press.

N. Asher and A. Lascarides. 2003. *Logics of Conversation*. Studies in Natural Language Processing. Cambridge University Press.

Jayeol Chun and Nianwen Xue. 2024. Uniform meaning representation parsing as a pipelined approach. In *Proceedings of TextGraphs-17: Graph-based Methods for Natural Language Processing*, pages 40–52, Bangkok, Thailand. Association for Computational Linguistics.

Jayeol Chun and Nianwen Xue. 2025. Modal dependency parsing via biaffine attention with self-loop. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 21226–21238, Vienna, Austria. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

DeepSeek-AI. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *Preprint*, arXiv:2501.12948.

DeepSeek-AI, Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang

Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Daya Guo, Dejian Yang, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, and 181 others. 2025. Deepseek-v3 technical report. *Preprint*, arXiv:2412.19437.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. Qlora: Efficient finetuning of quantized llms. In *Advances in Neural Information Processing Systems*, volume 36, pages 10088–10115. Curran Associates, Inc.

Mona Diab, Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaran, and Weiwei Guo. 2009. Committed belief annotation and tagging. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, pages 68–73, Suntec, Singapore. Association for Computational Linguistics.

Kenton Lee, Yoav Artzi, Yejin Choi, and Luke Zettlemoyer. 2015. Event detection and factuality assessment with non-expert supervision. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1643–1648, Lisbon, Portugal. Association for Computational Linguistics.

Zhifu Liu and Nianwen Xue. 2023. A dependency structure annotation for modality in chinese news articles. In *Chinese Lexical Semantics*, pages 143–157, Cham. Springer Nature Switzerland.

Amnon Lotan, Asher Stern, and Ido Dagan. 2013. TruthTeller: Annotating predicate truth. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 752–757, Atlanta, Georgia. Association for Computational Linguistics.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, Maryland. Association for Computational Linguistics.

Rowan Nairn, Cleo Condoravdi, and Lauri Karttunen. 2006. Computing relative polarity for textual inference. In *Proceedings of the Fifth International Workshop on Inference in Computational Semantics (ICoS-5)*.

F. R. Palmer. 2001. *Mood and Modality*, 2 edition. Cambridge Textbooks in Linguistics. Cambridge University Press.

Paul Portner. 2009. *Modality*.

Amir Pouran Ben Veyseh, Thien Huu Nguyen, and Dejing Dou. 2019. Graph based neural networks for event factuality prediction using syntactic and semantic structures. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4393–4399, Florence, Italy. Association for Computational Linguistics.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Zhong Qian, Peifeng Li, Yue Zhang, Guodong Zhou, and Qiaoming Zhu. 2018. Event factuality identification via generative adversarial networks with auxiliary classification. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 4293–4300. International Joint Conferences on Artificial Intelligence Organization.

Qwen, :, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, and 25 others. 2025. Qwen2.5 technical report. *Preprint*, arXiv:2412.15115.

Rachel Rudinger, Aaron Steven White, and Benjamin Van Durme. 2018. Neural models of factuality. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 731–744, New Orleans, Louisiana. Association for Computational Linguistics.

Roser Saurí and James Pustejovsky. 2009. Factbank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3):227–268.

Roser Saurí and James Pustejovsky. 2012. Are you sure that this happened? assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299.

E. Semino and M. Short. 2011. *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing*. Routledge advances in corpus linguistics. Routledge.

Gabriel Stanovsky, Judith Eckle-Kohler, Yevgeniy Puzikov, Ido Dagan, and Iryna Gurevych. 2017. Integrating deep linguistic features in factuality prediction over unified datasets. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 352–357, Vancouver, Canada. Association for Computational Linguistics.

Jens E. L. Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O'Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, Jan Hajič, James H. Martin, Stephan Oepen, Martha Palmer, James Pustejovsky, Rosa Vallejos, and Nianwen Xue. 2021. Designing a uniform meaning representation for natural language processing. *KI - Künstliche Intelligenz*, 35(3):343–360.

Meagan Vigus, Jens E. L. Van Gysel, and William Croft. 2019. A dependency structure annotation for modality. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 182–198, Florence, Italy. Association for Computational Linguistics.

Jiarui Yao, Haoling Qiu, Jin Zhao, Bonan Min, and Nianwen Xue. 2021. Factuality assessment as modal dependency parsing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1540–1550, Online. Association for Computational Linguistics.

Jiarui Yao, Nianwen Xue, and Bonan Min. 2022. Modal dependency parsing via language model priming. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2913–2919, Seattle, United States. Association for Computational Linguistics.

A Experimental Details

All experiments are conducted on a single NVIDIA RTX A6000 GPU. Under the QLoRA configuration, the number of trainable parameters is 161,480,704, corresponding to approximately 2.1% of the total 7,777,097,216 model parameters.

A.1 Input Representation

Input documents are linearized as token-indexed text spans, where each sentence is prefixed with its document-level token range:

> [0, 9) The witness claimed she saw the suspect yesterday .
> [9, 14) Police are investigating the case .

For exceptionally long sentences, we apply additional segmentation into smaller units to improve stability and reduce sequence length variability.

A.2 Output Representation

The output is a structured JSON encoding document-level semantic dependencies, where all entities are assigned unique identifiers and spans are represented using global token indices. A sample output JSON is shown in Figure 6.

Category	Hyperparameter	Default Value
Base Model	Base LLM	DeepSeek-R1-Distill-Qwen-7B
	Precision	4-bit QLoRA
Optimization	Optimizer	AdamW
	Learning rate	2×10^{-5}
	Weight decay	0.01
	Gradient clipping	1.0
Training	Batch size (per device)	4
	Gradient accumulation	4
	Number of epochs	10
	Warmup ratio	0.03
	LR scheduler	Linear
LoRA / QLoRA	LoRA rank	64
	LoRA alpha	16
	LoRA dropout	0.1
	Target modules	all linear layers
Decoding	Decoding strategy	Greedy
	Temperature	0.0
	Top P	1.0
Reproducibility	Aggregation	Single-Best
	Seed	42

Table 4: Hyperparameters for SFT

Output (simplified)


```
{
  "events": [
    {
        "id": "e0",
        "span": [2, 3]
    },
    ...
  ],
  "targets": [
    {
        "id": "t0",
        "src": [1, 2],
        "cue": [2, 3],
        "scope": [3, 9]
    },
    ...
  ],
  "edges": [
    {
        "child": "e0",
        "head": "AUTHOR",
        "pol": "+",
        "str": "full"
    },
    ...
  ]
}
```


Figure 6: Example JSON output