GLaRef@CRAC2025: Should We Transform Coreference Resolution into a Text Generation Task?

Olga Seminck, **Antoine Bourgois**, Yoann Dupont, Mathieu Dehouck, Marine Delaborde

Lattice (CNRS – École Normale Supérieure – Université Sorbonne Nouvelle), Paris, France

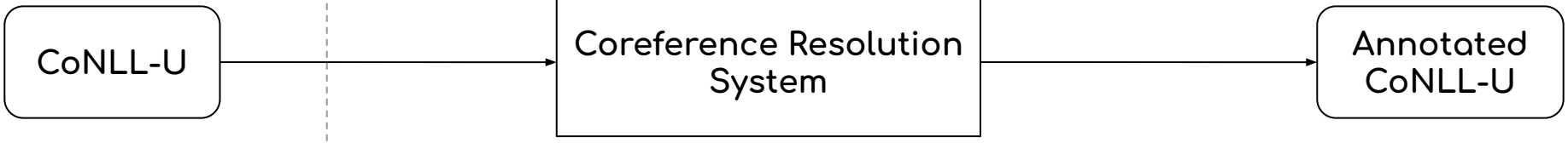
Eighth Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2025)

Shared Task Overview

- CorefUD 1.3
- 17 Languages :
  Ancient Greek, Biblical Hebrew, Catalan, Czech, English, French, German, Hindi, Hungarian, Korean, Lithuanian, Norwegian, Old Church Slavonic, Polish, Russian, Spanish, Turkish
- 22 Datasets
  - ↳ Documents
    - ↳ Sentences

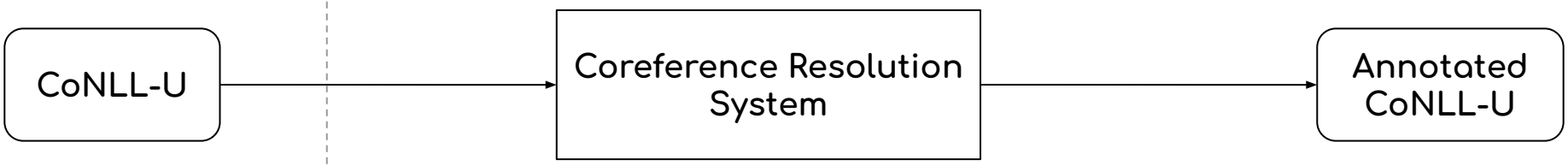
Shared Task Overview: Overview

22 Datasets



Shared Task Overview: Overview

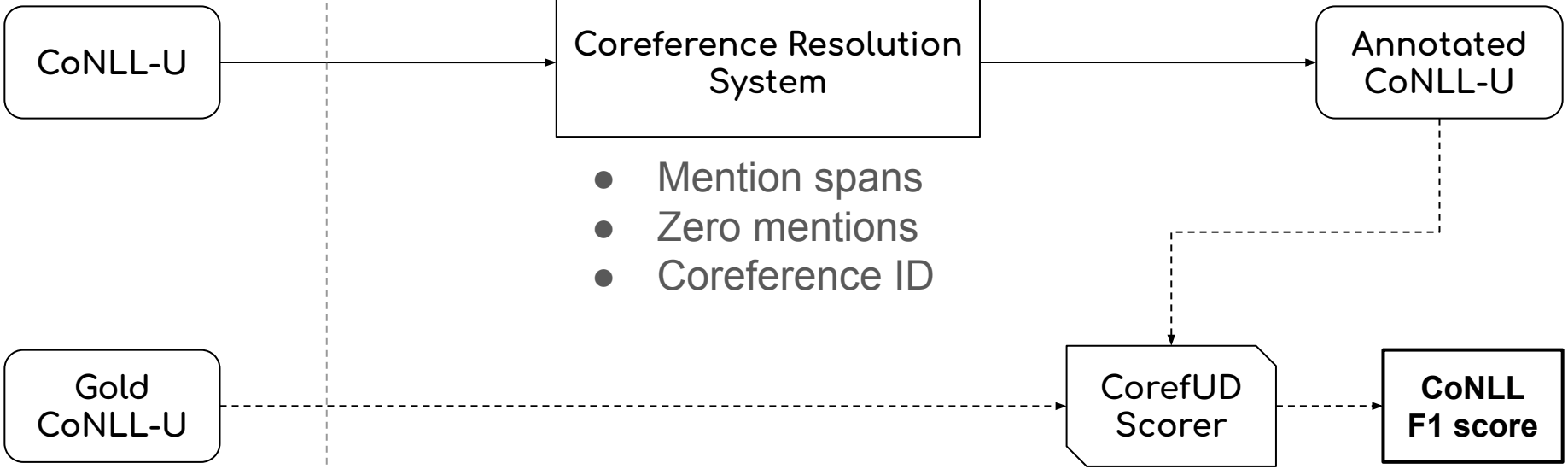
22 Datasets



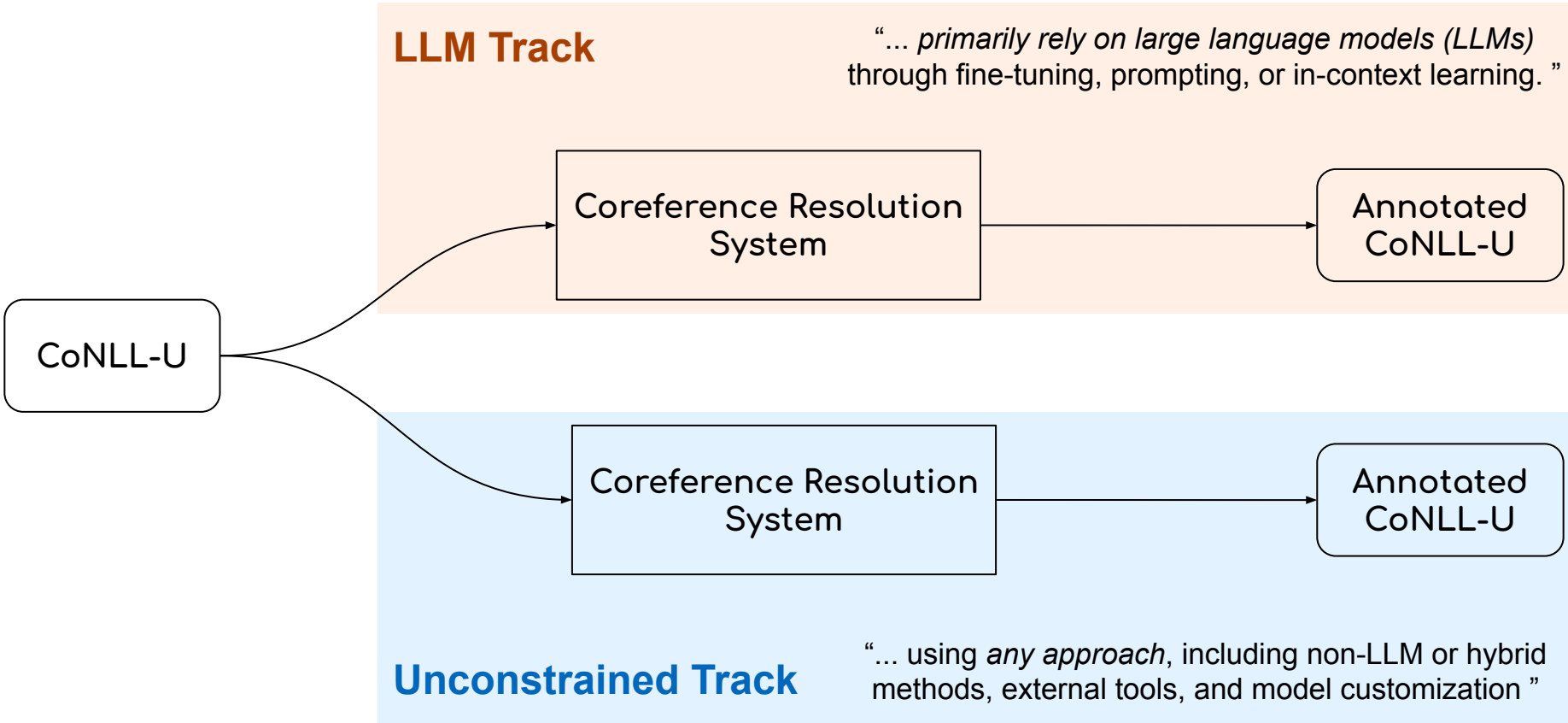
- Mention spans
- Zero mentions
- Coreference ID

Shared Task Overview: Evaluation

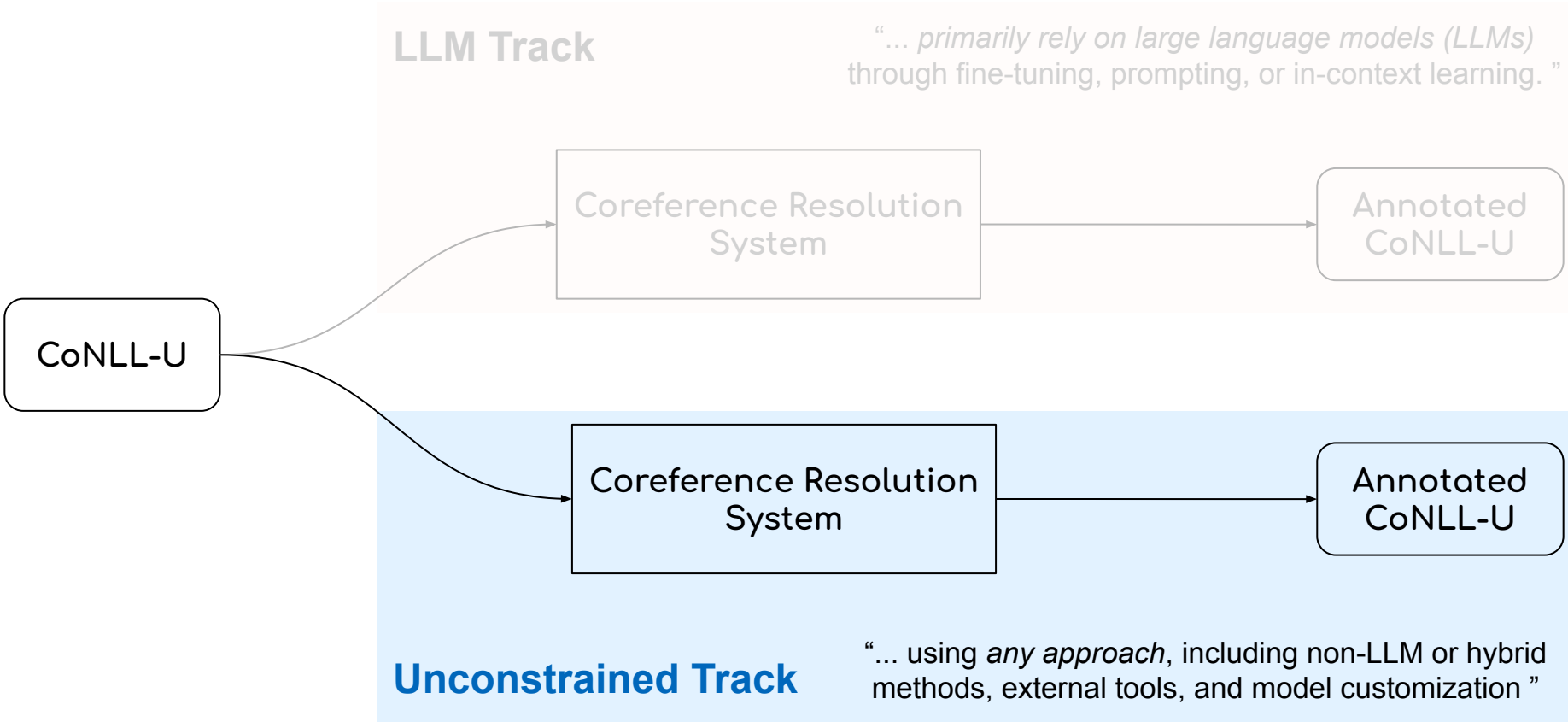
22 Datasets



Unconstrained & LLM Tracks

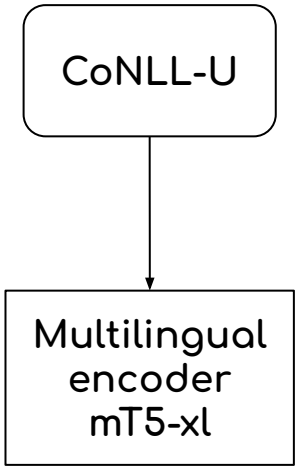


Unconstrained Track



Unconstrained Submission: Multistage Pipeline

Unconstrained Submission: Multistage Pipeline



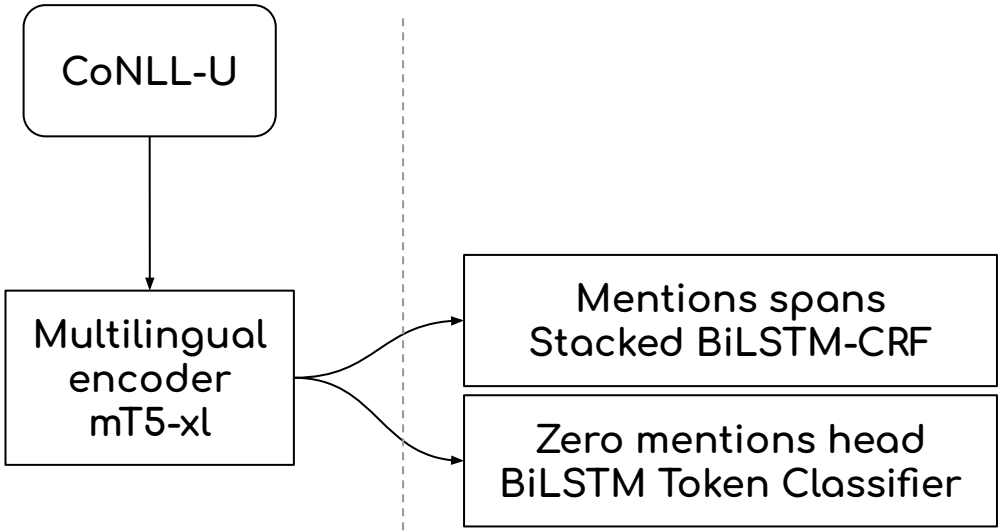
Token Embeddings

Mentions Detection

Mention Pairs

Clustering

Unconstrained Submission: Multistage Pipeline



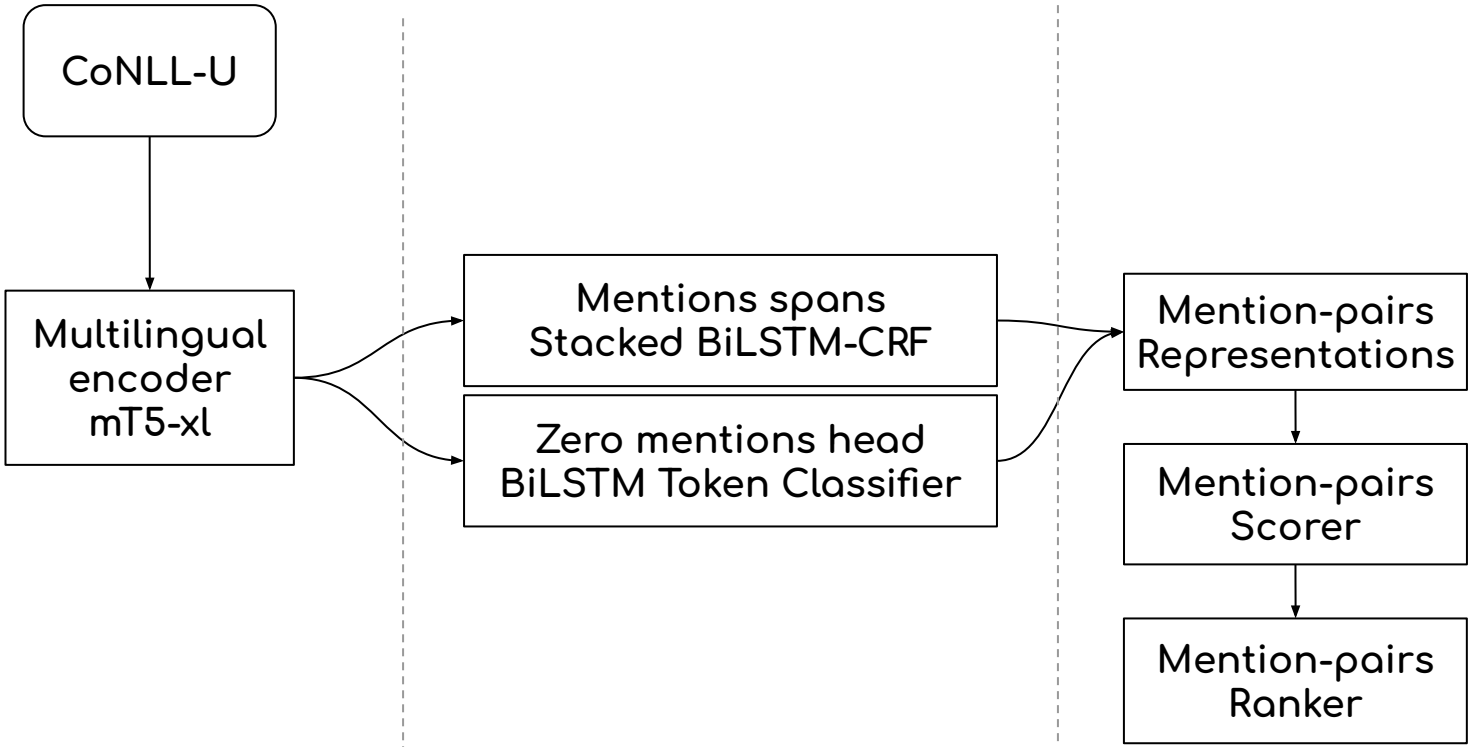
Token Embeddings

Mentions Detection

Mention Pairs

Clustering

Unconstrained Submission: Multistage Pipeline



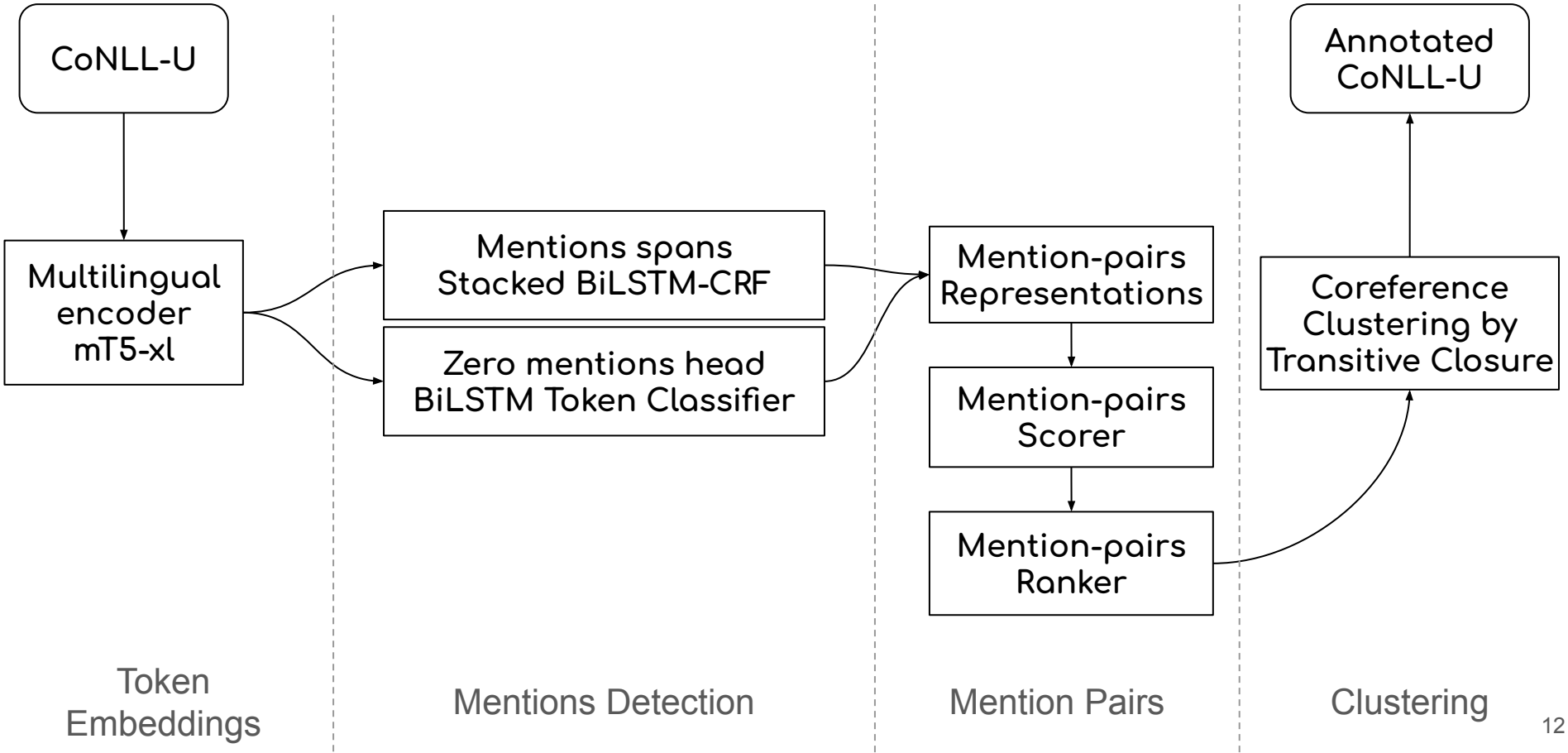
Token Embeddings

Mentions Detection

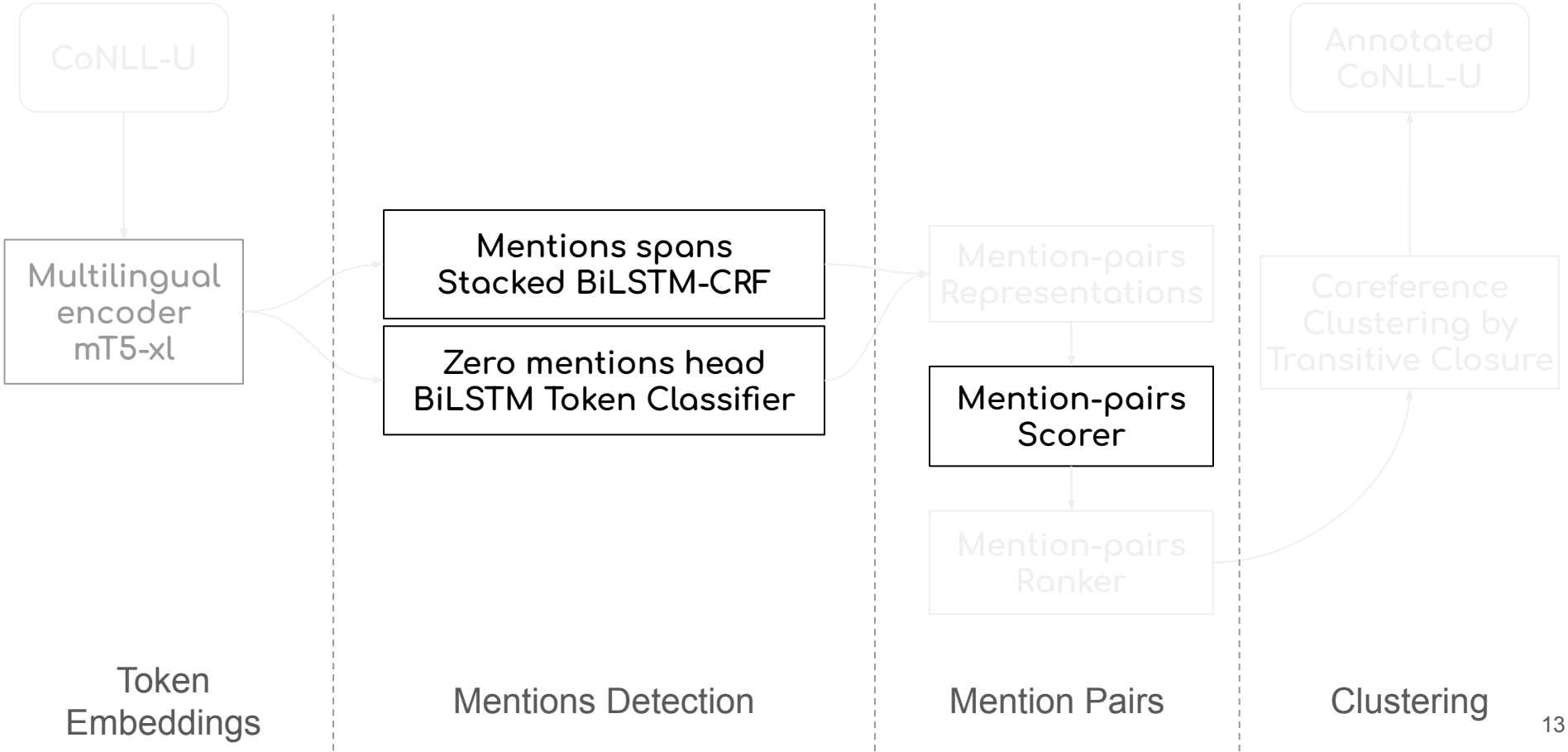
Mention Pairs

Clustering

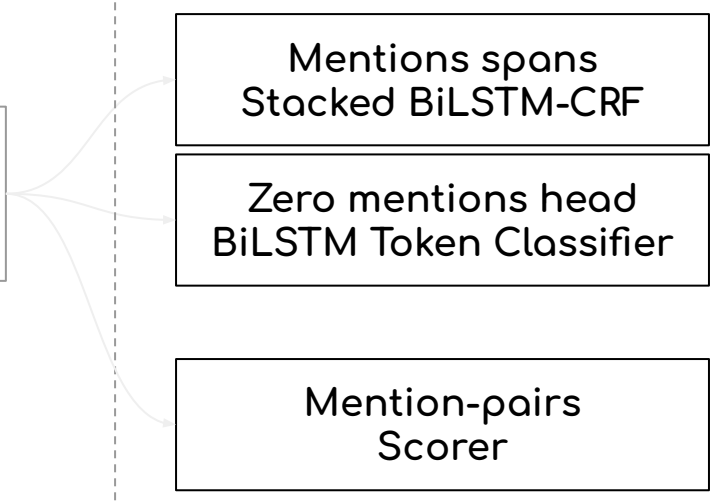
Unconstrained Submission: Multistage Pipeline



Unconstrained Submission: Trained Modules



Unconstrained Submission: Training Resources



Token Embeddings

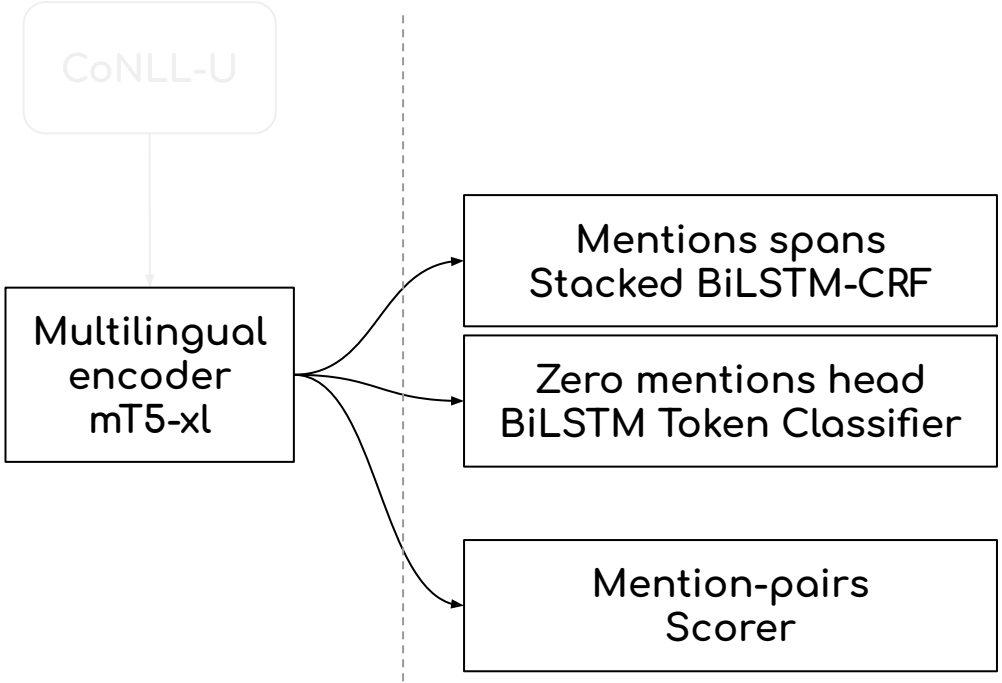
	GPU Peak VRAM	Training Time
Mentions spans Stacked BiLSTM-CRF	3.8 GiB	< 5 hours
Zero mentions head BiLSTM Token Classifier	1.7 GiB	< 4 hours
Mention-pairs Scorer	1.7 GiB	< 3 hours

Unconstrained Submission: Training Resources

Token Embeddings

	GPU Peak VRAM	Training Time
Mentions spans Stacked BiLSTM-CRF	3.8 GiB	< 5 hours
Zero mentions head BiLSTM Token Classifier	1.7 GiB	< 4 hours
Mention-pairs Scorer	1.7 GiB	< 3 hours
	7.2 GiB	**< 5 hours**

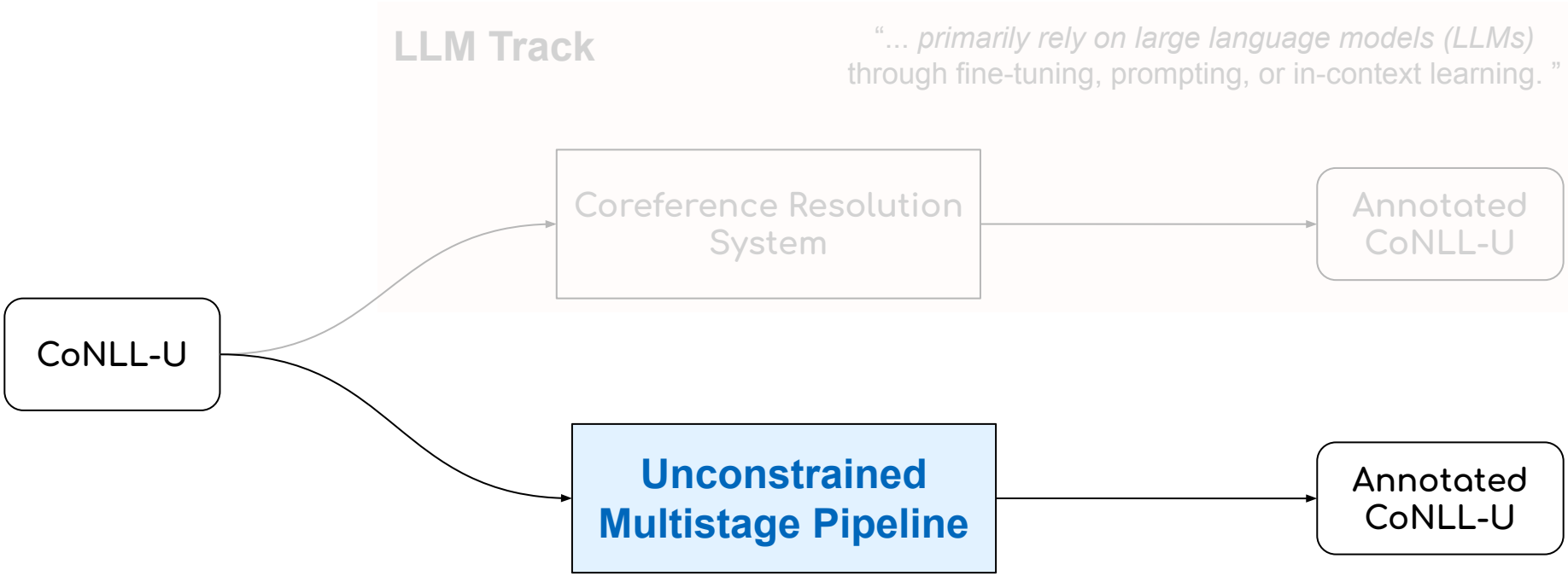
Unconstrained Submission: Training Resources



	GPU Peak VRAM	Training Time
Mentions spans Stacked BiLSTM-CRF	3.8 GiB	< 5 hours
Zero mentions head BiLSTM Token Classifier	1.7 GiB	< 4 hours
Mention-pairs Scorer	1.7 GiB	< 3 hours
	7.2 GiB	**< 5 hours**

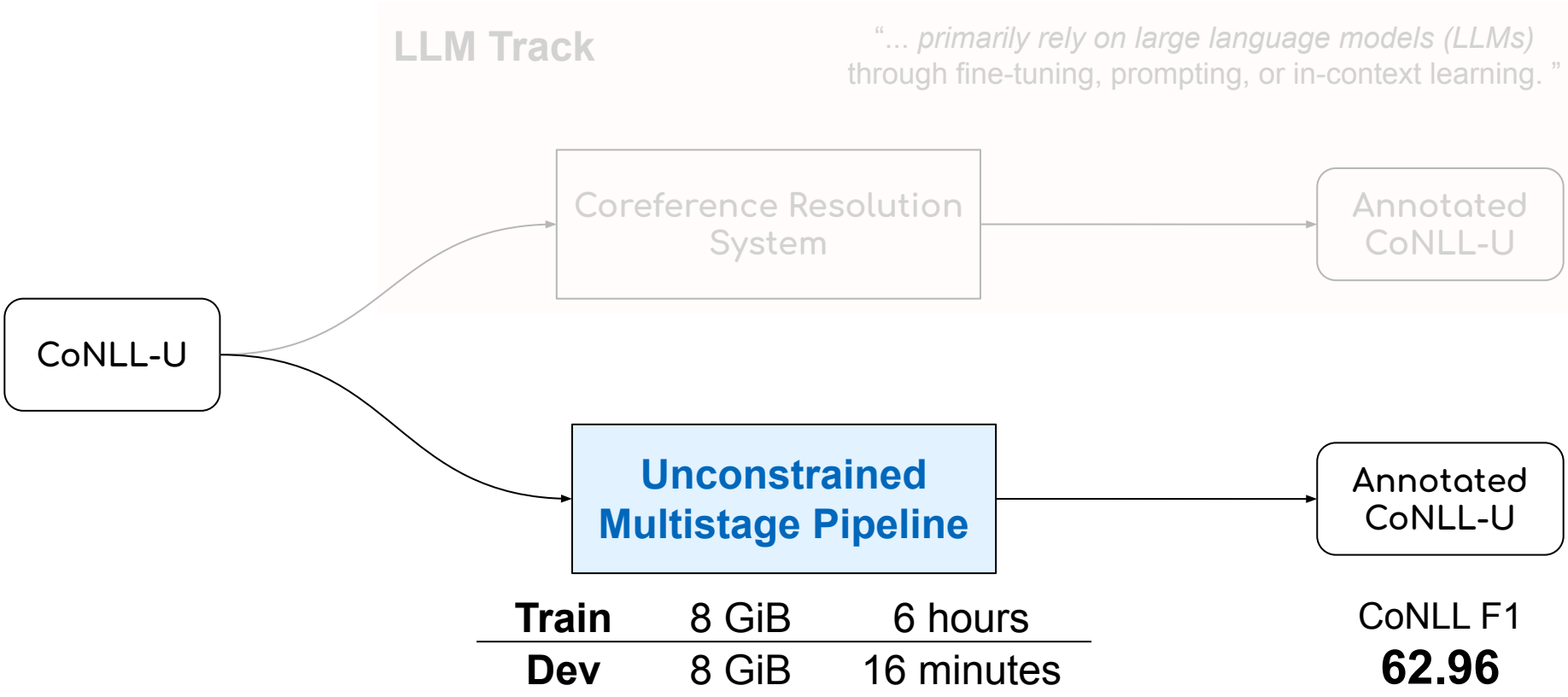
- **8 GiB**
- **< 1 hour**

Unconstrained Submission: Training Resources

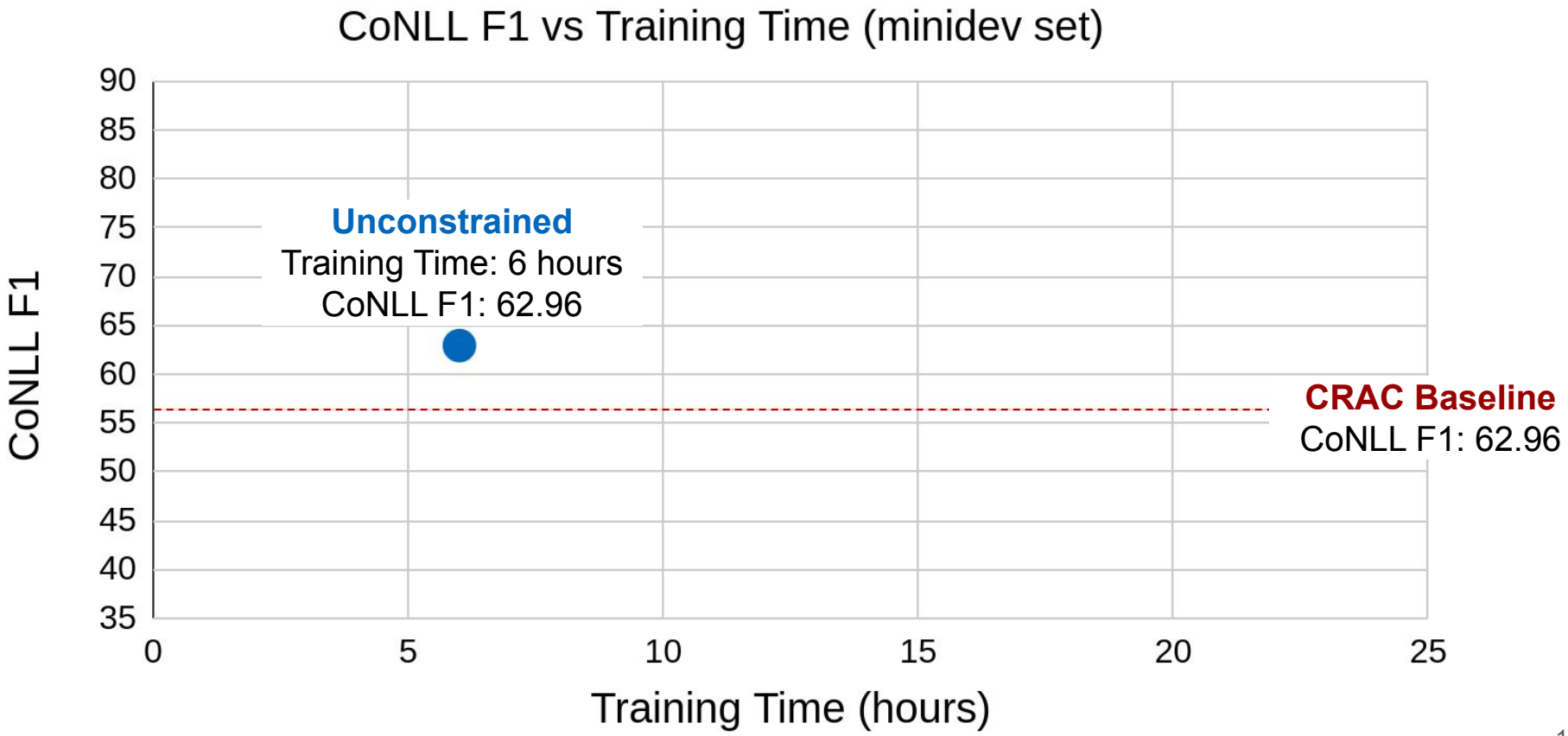


Train	8 GiB	6 hours
Dev	8 GiB	16 minutes

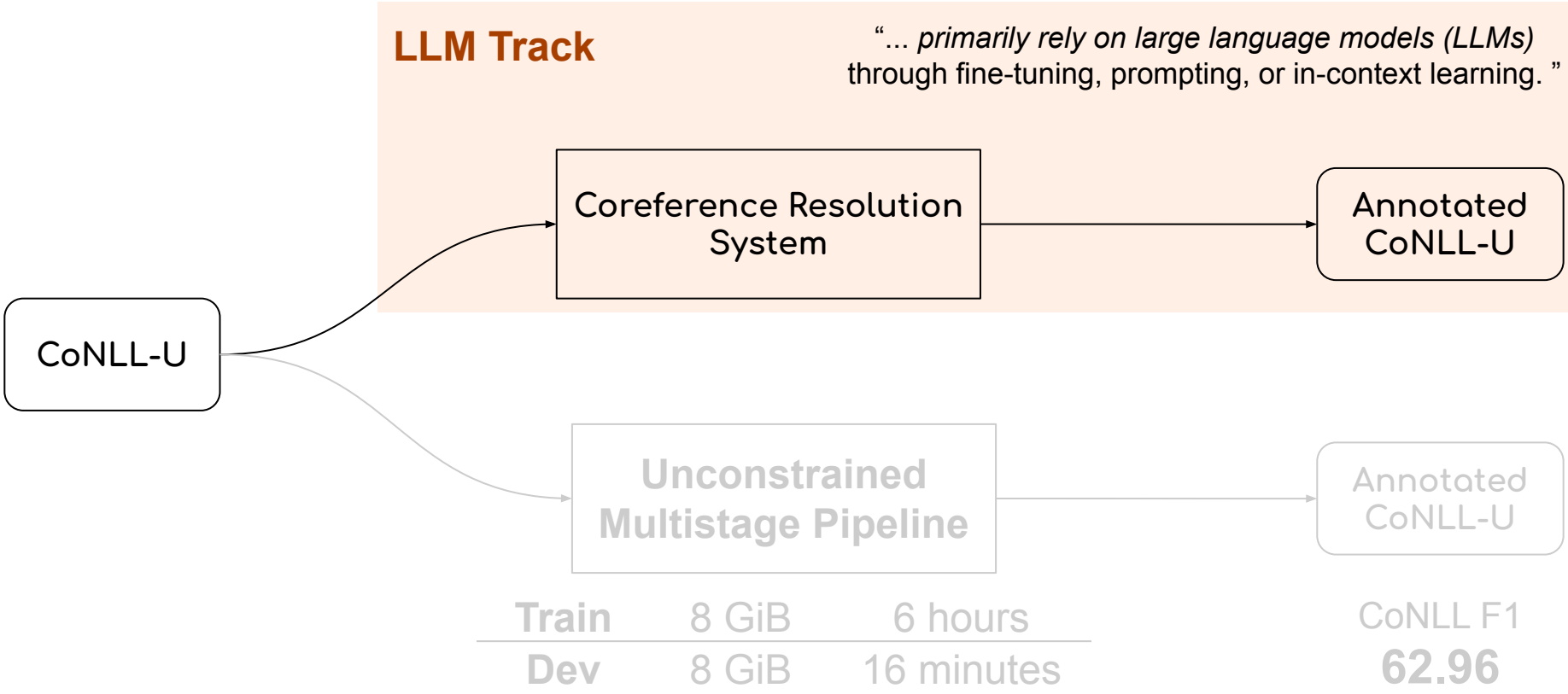
Unconstrained Submission: Results



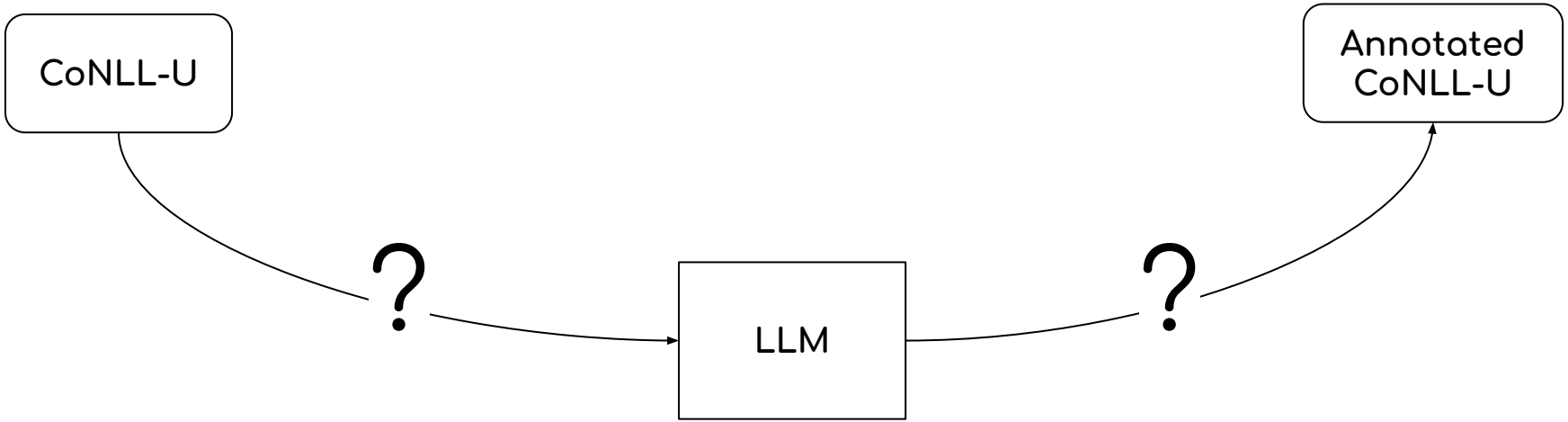
Comparison of Models



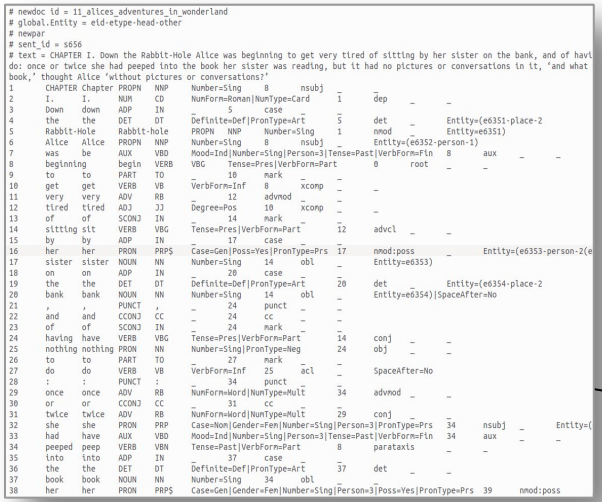
LLM Track


LLM Track
"... primarily rely on large language models (LLMs) through fine-tuning, prompting, or in-context learning. "
Coreference Resolution System
Annotated CoNLL-U
CoNLL-U
Unconstrained Multistage Pipeline
Annotated CoNLL-U
Train 8 GiB 6 hours
Dev 8 GiB 16 minutes
CoNLL F1
62.96

LLM Track



LLM Track: Provided Conversion Scripts



```
# newdoc id = 11_alices_adventures_in_wonderland
# global.Entity = eid-etype-head-other
# newpar
# sent_id = s656
# text = CHAPTER I. Down the Rabbit-Hole Alice was beginning to get very tired of sitting by her sister on the bank, and of havi
do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what
book,' thought Alice 'without pictures or conversations?'
1       CHAPTER Chapter PROPN   NNP     Number=Sing     8       nsubj   _       _
2       I.      I.      NUM     CD      NumForm=Roman|NumType=Card      1       dep     _       _
3       Down    down    ADP     IN      _       5       case    _       _
4       the     the     DET     DT      Definite=Def|PronType=Art       5       det     _       Entity=(e6351-place-2
5       Rabbit-Hole     Rabbit-hole     PROPN   NNP     Number=Sing     1       nmod    _       Entity=e6351)
6       Alice   Alice   PROPN   NNP     Number=Sing     8       nsubj   _       Entity=(e6352-person-1)
7       was     be      AUX     VBD     Mood=Ind|Number=Sing|Person=3|Tense=Past|VerbForm=Fin   8       aux     _       _
8       beginning       begin   VERB    VBG     Tense=Pres|VerbForm=Part        0       root    _       _
9       to      to      PART    TO      _       10      mark    _       _
10      get     get     VERB    VB      VerbForm=Inf    8       xcomp   _       _
11      very    very    ADV     RB      _       12      advmod  _       _
12      tired   tired   ADJ     JJ      Degree=Pos      10      xcomp   _       _
13      of      of      SCONJ   IN      _       14      mark    _       _
14      sitting sit     VERB    VBG     Tense=Pres|VerbForm=Part        12      advcl   _       _
15      by      by      ADP     IN      _       17      case    _       _
16      her     her     PRON    PRP$    Case=Gen|Poss=Yes|PronType=Prs  17      nmod:poss       _       Entity=(e6353-person-2(e
17      sister  sister  NOUN    NN      Number=Sing     14      obl     _       Entity=e6353)
18      on      on      ADP     IN      _       20      case    _       _
19      the     the     DET     DT      Definite=Def|PronType=Art       20      det     _       Entity=(e6354-place-2
20      bank    bank    NOUN    NN      Number=Sing     14      obl     _       Entity=e6354)|SpaceAfter=No
21      ,       ,       PUNCT   ,       _       24      punct   _       _
22      and     and     CCONJ   CC      _       24      cc      _       _
23      of      of      SCONJ   IN      _       24      mark    _       _
24      having  have    VERB    VBG     Tense=Pres|VerbForm=Part        14      conj    _       _
25      nothing nothing PRON    NN      Number=Sing|PronType=Neg        24      obj     _       _
26      to      to      PART    TO      _       27      mark    _       _
27      do      do      VERB    VB      VerbForm=Inf    25      acl     _       SpaceAfter=No
28      :       :       PUNCT   :       _       34      punct   _       _
29      once    once    ADV     RB      NumForm=Word|NumType=Mult       34      advmod  _       _
30      or      or      CCONJ   CC      _       31      cc      _       _
31      twice   twice   ADV     RB      NumForm=Word|NumType=Mult       29      conj    _       _
32      she     she     PRON    PRP     Case=Nom|Gender=Fem|Number=Sing|Person=3|PronType=Prs   34      nsubj   _       Entity=(
33      had     have    AUX     VBD     Mood=Ind|Number=Sing|Person=3|Tense=Past|VerbForm=Fin   34      aux     _       _
34      peeped  peep    VERB    VBN     Tense=Past|VerbForm=Part        8       parataxis       _       _
35      into    into    ADP     IN      _       37      case    _       _
36      the     the     DET     DT      Definite=Def|PronType=Art       37      det     _       _
37      book    book    NOUN    NN      Number=Sing     34      obl     _       _
38      her     her     PRON    PRP$    Case=Gen|Gender=Fem|Number=Sing|Person=3|Poss=Yes|PronType=Prs   39      nmod:poss
```

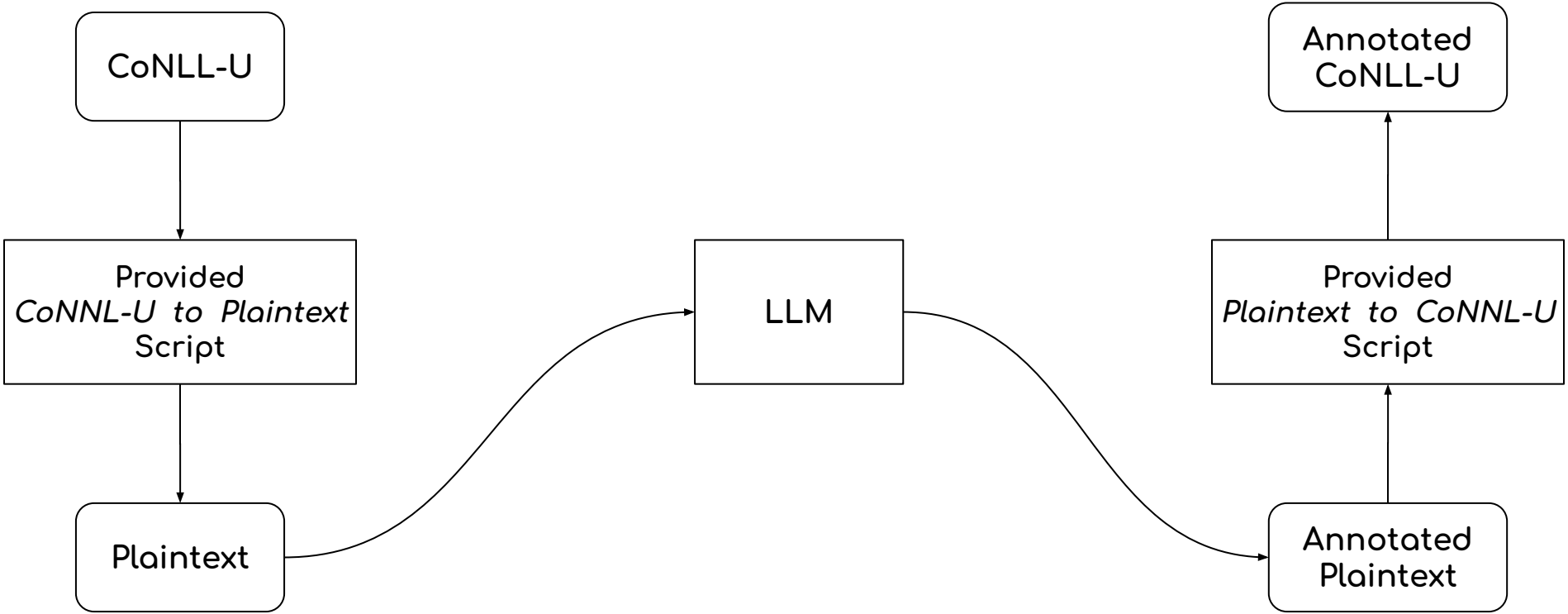
LLM

?

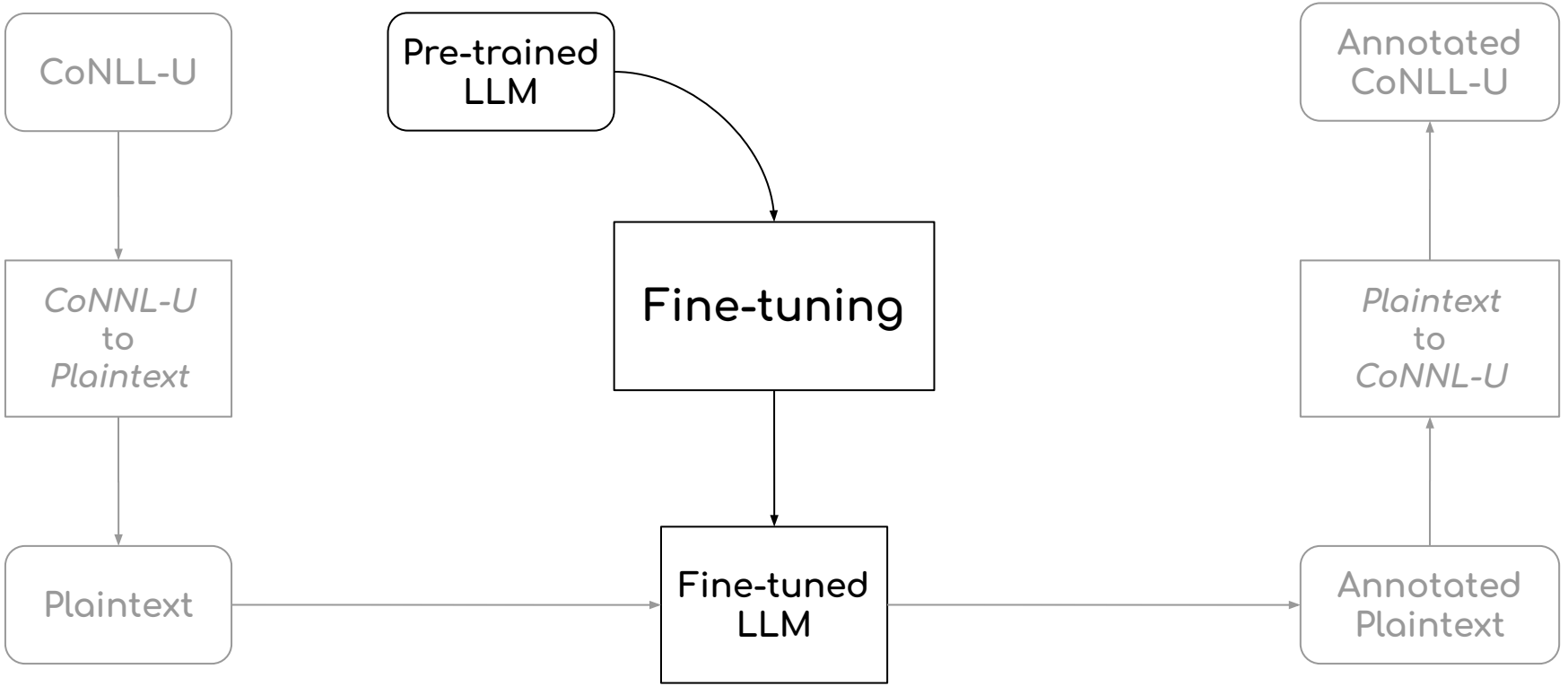
Annotated CoNLL-U

LLMs are not designed to process CoNLLU-formatted data

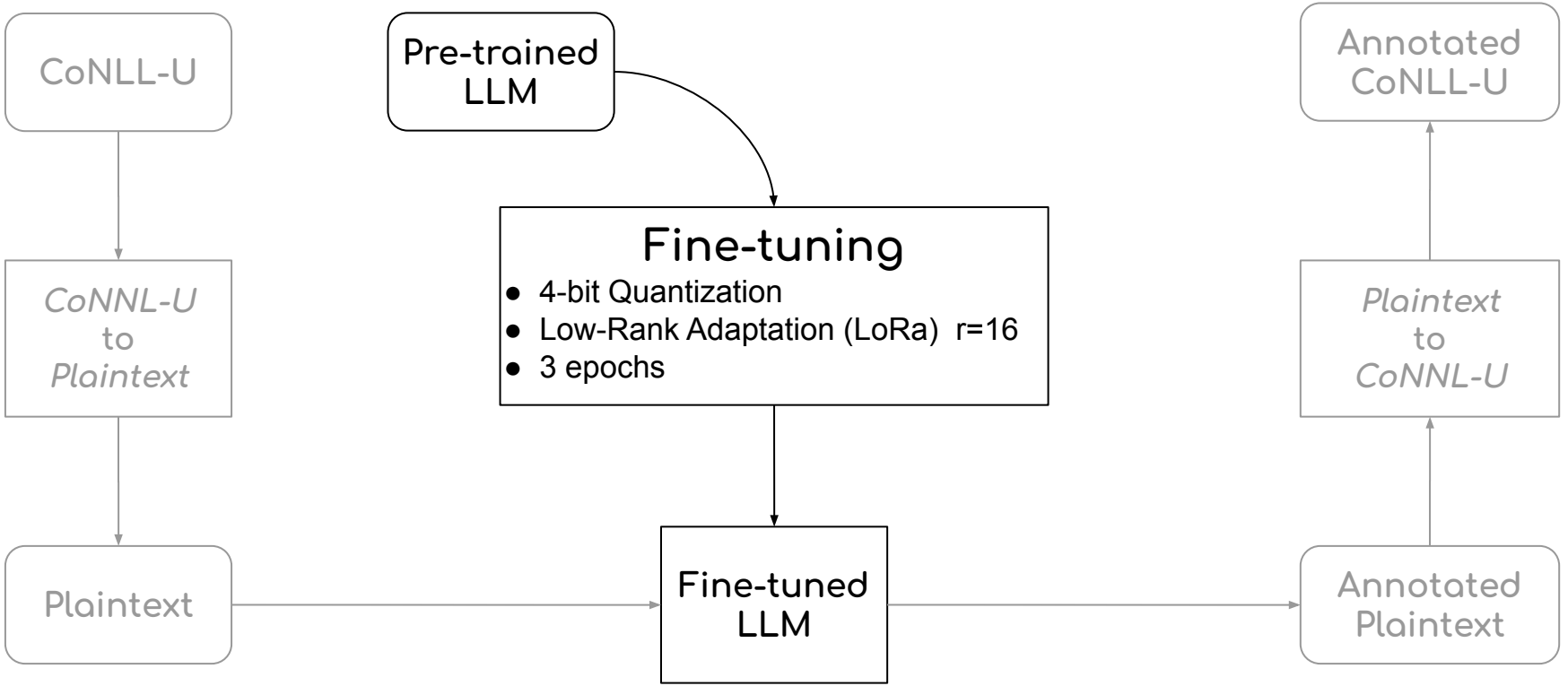
LLM Track: Provided Conversion Scripts



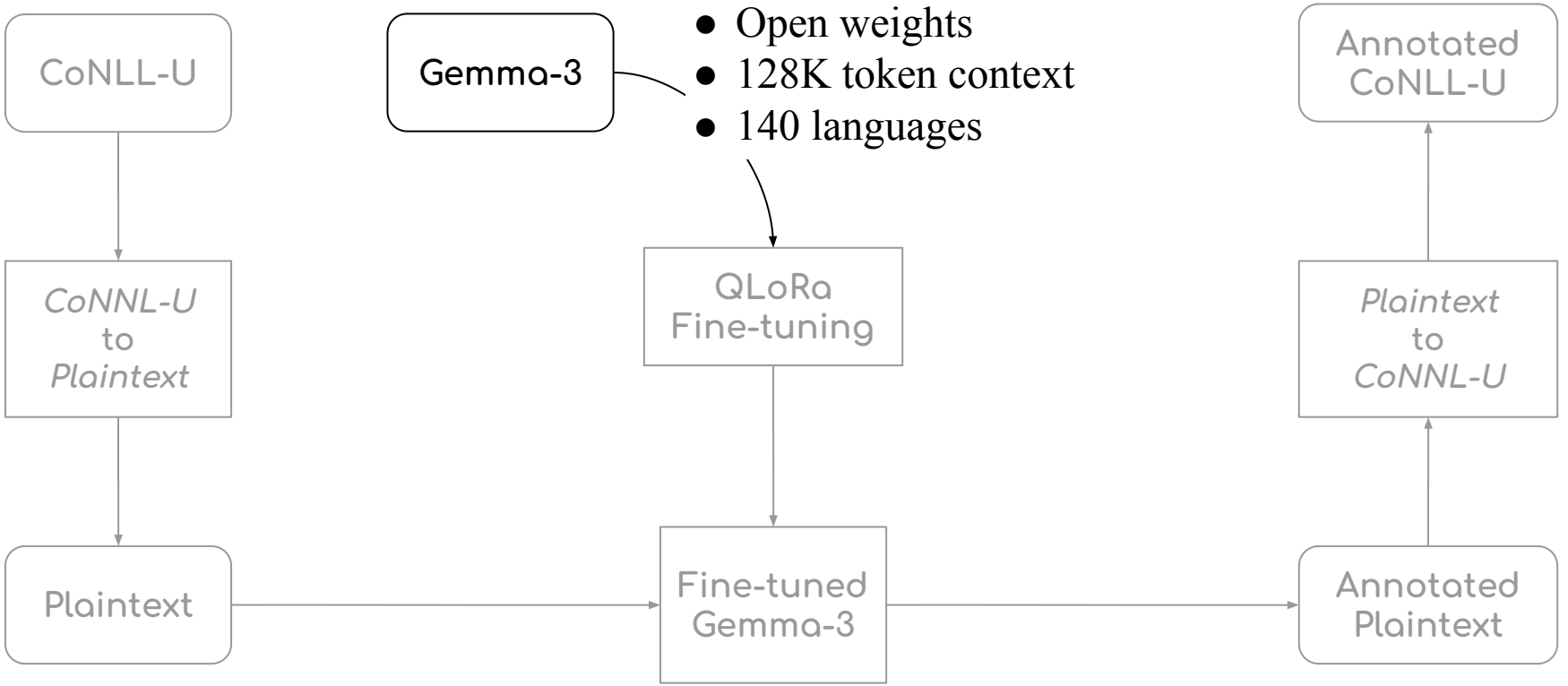
LLM Fine-tuning



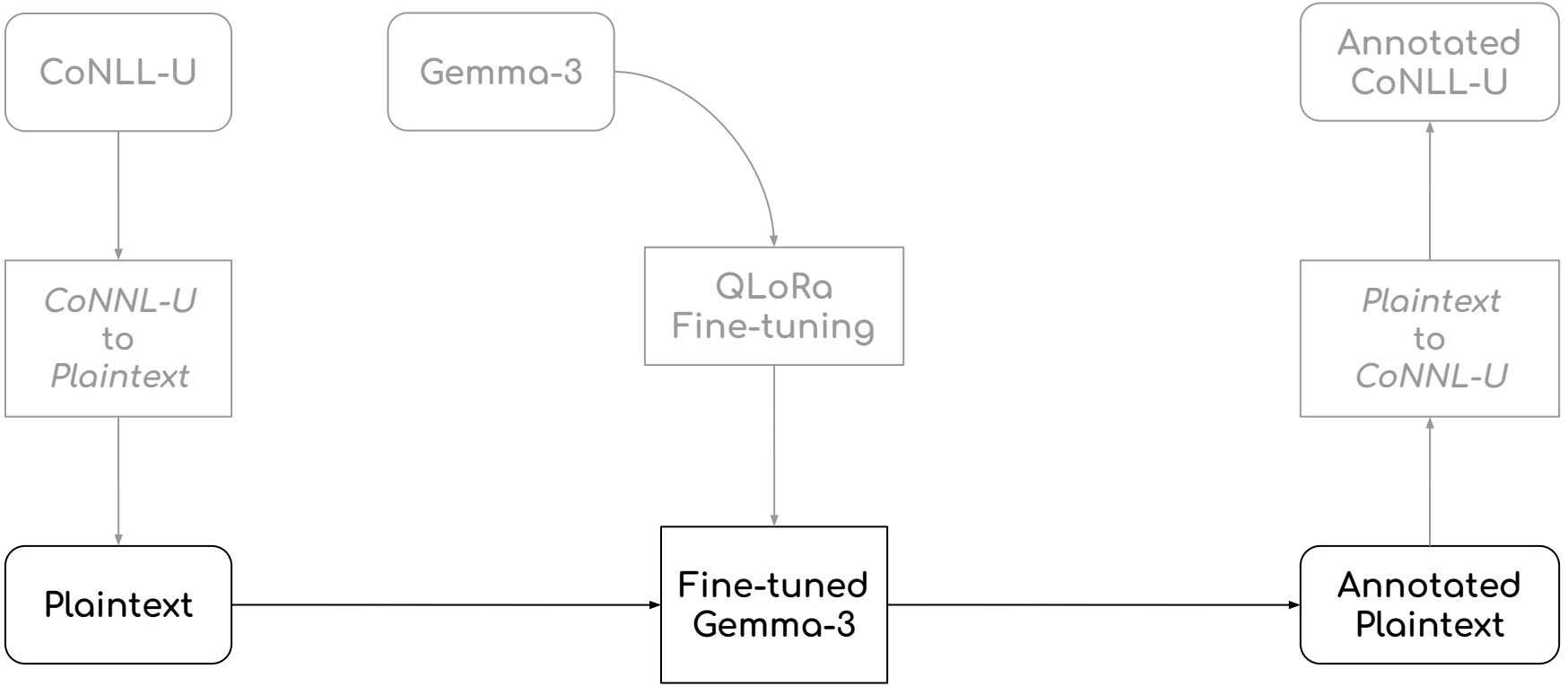
LLM Fine-tuning



Pre-trained LLM Choice: Gemma-3 instruction-tuned (IT)



Input Formatting

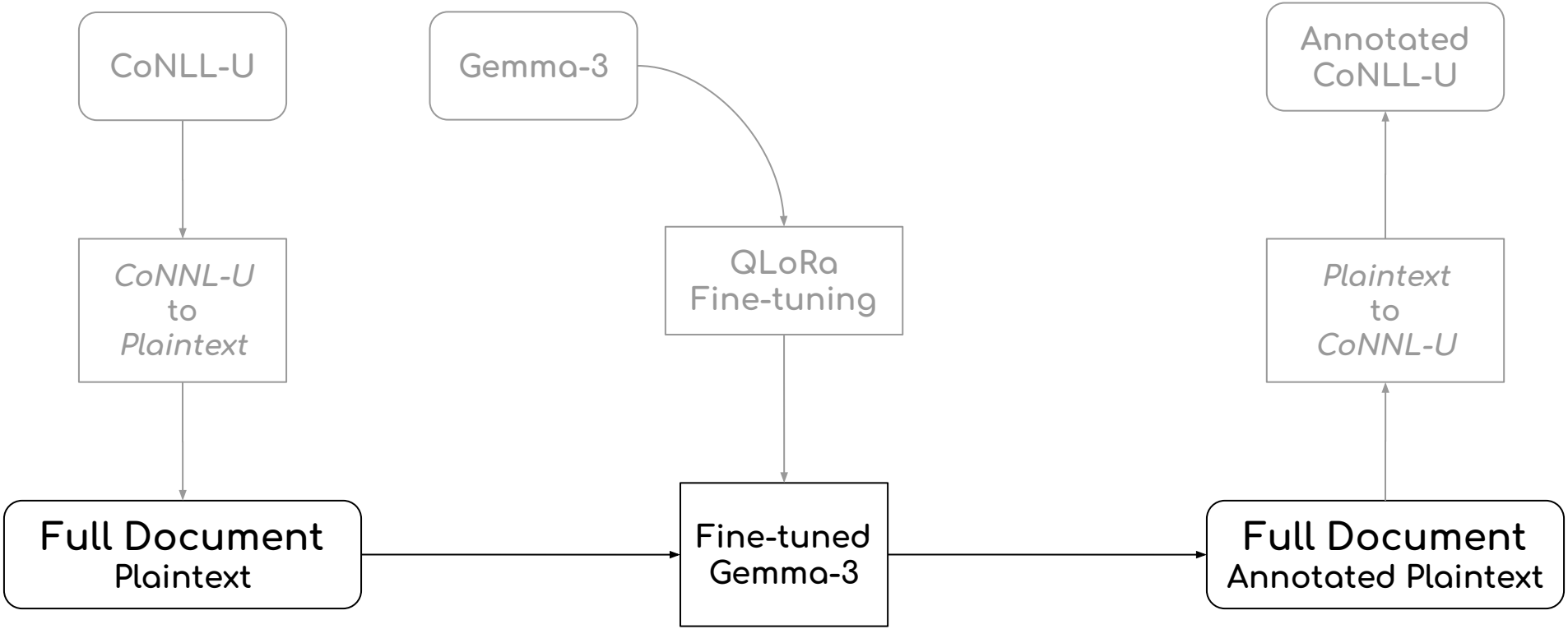


Gemma-3 Prompt Template

SYSTEM INSTRUCTION TEXT INPUT EXPECTED MODEL OUTPUT

```
<start_of_turn>user
You are a linguist, expert in anaphora and coreference resolution.
Annotate in the input sentences which nouns, pronouns and other expressions
refer to the same entity.
Do only insert annotations. Do not insert extra linguistic material, nor
punctuation markers and do not delete elements from the input texts.

Input: *PLAINTEXT*
<end_of_turn>

<start_of_turn>model
*COREFERENCE ANNOTATED PLAINTEXT*
<end_of_turn>
```

Full Document Annotation

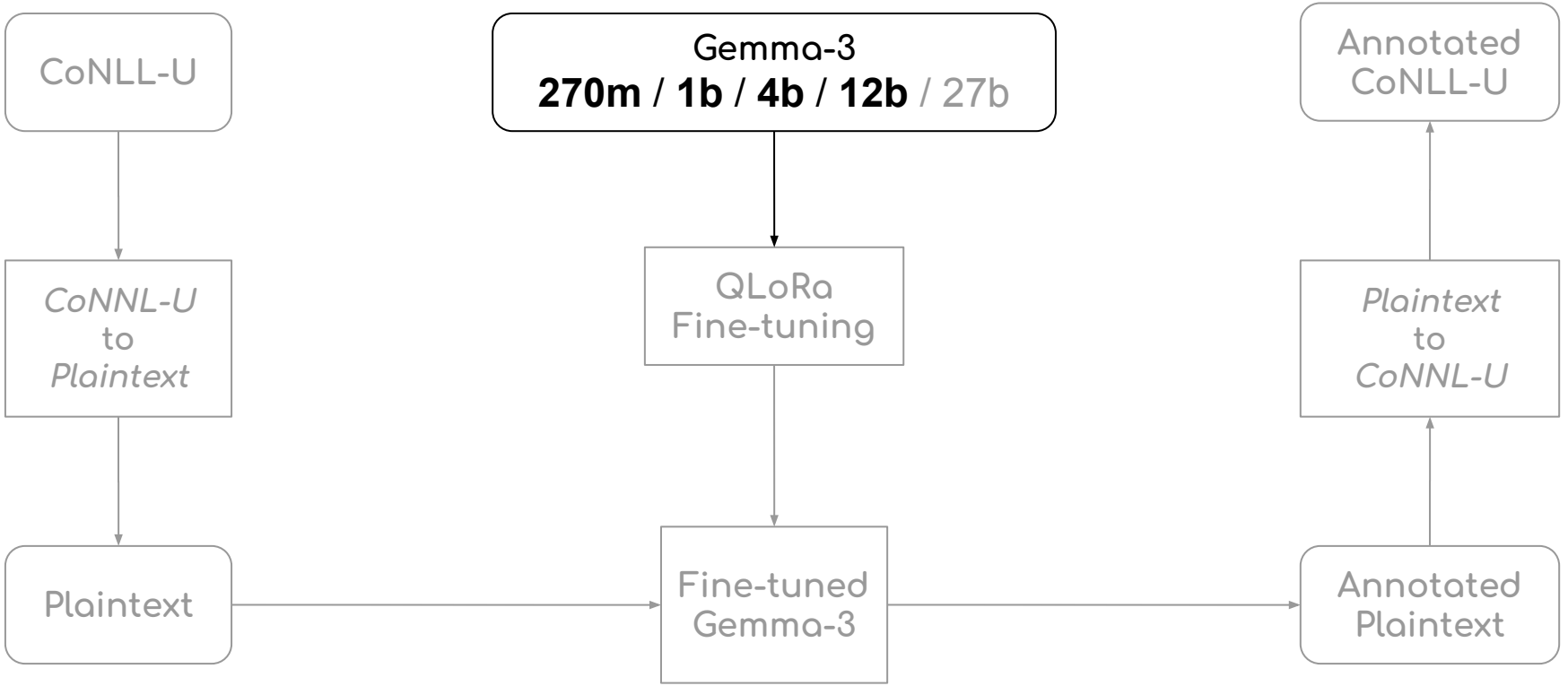


Gemma-3: Prompt Template

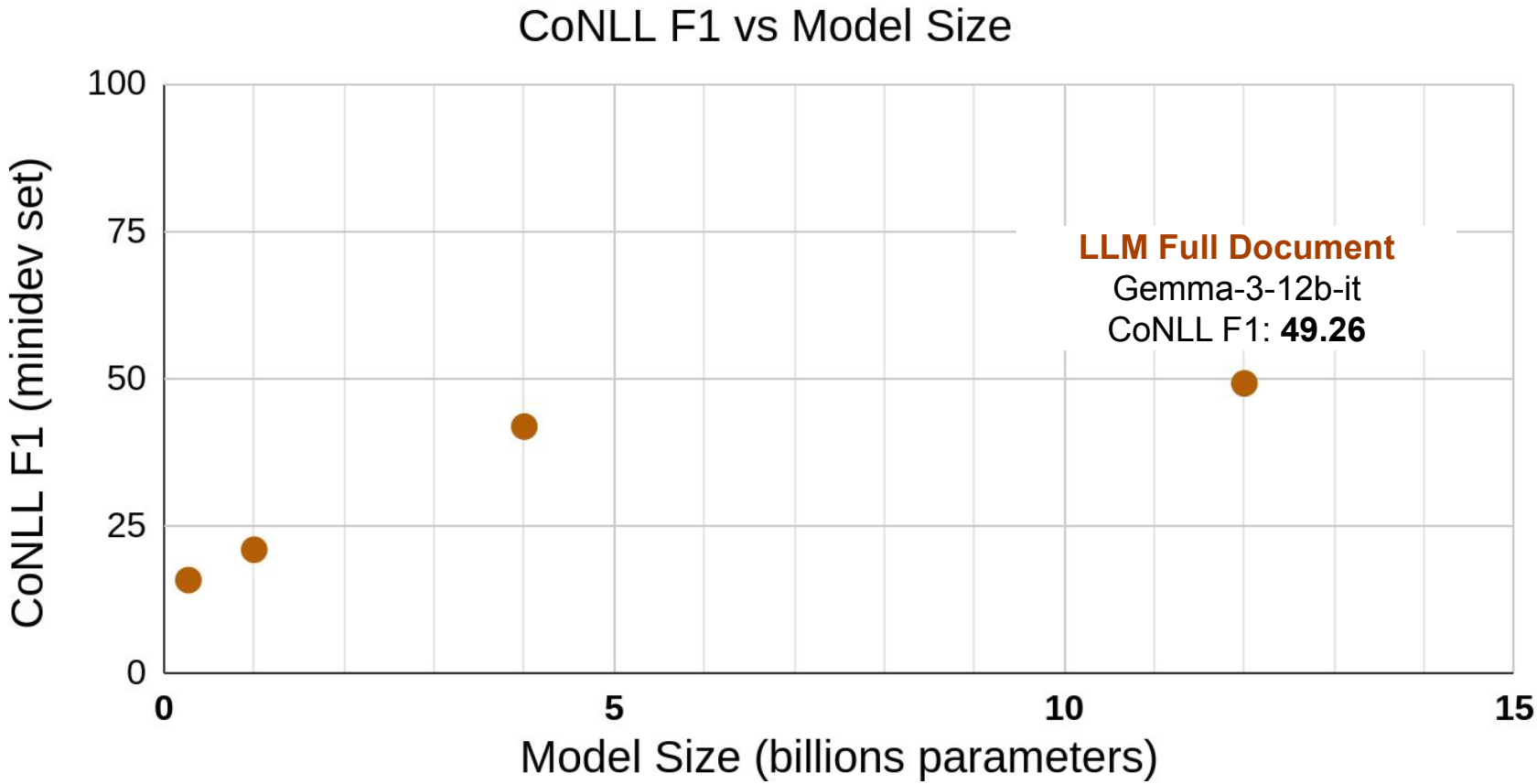
SYSTEM INSTRUCTION TEXT INPUT EXPECTED MODEL OUTPUT

```
<start_of_turn>user
You are a linguist, expert in anaphora and coreference resolution.
Annotate in the input sentences which nouns, pronouns and other expressions
refer to the same entity.
Do only insert annotations. Do not insert extra linguistic material, nor
punctuation markers and do not delete elements from the input texts.

Input: *PLAINTEXT*
<end_of_turn>

<start_of_turn>model
*COREFERENCE ANNOTATED SENTENCE BATCH*
<end_of_turn>
```

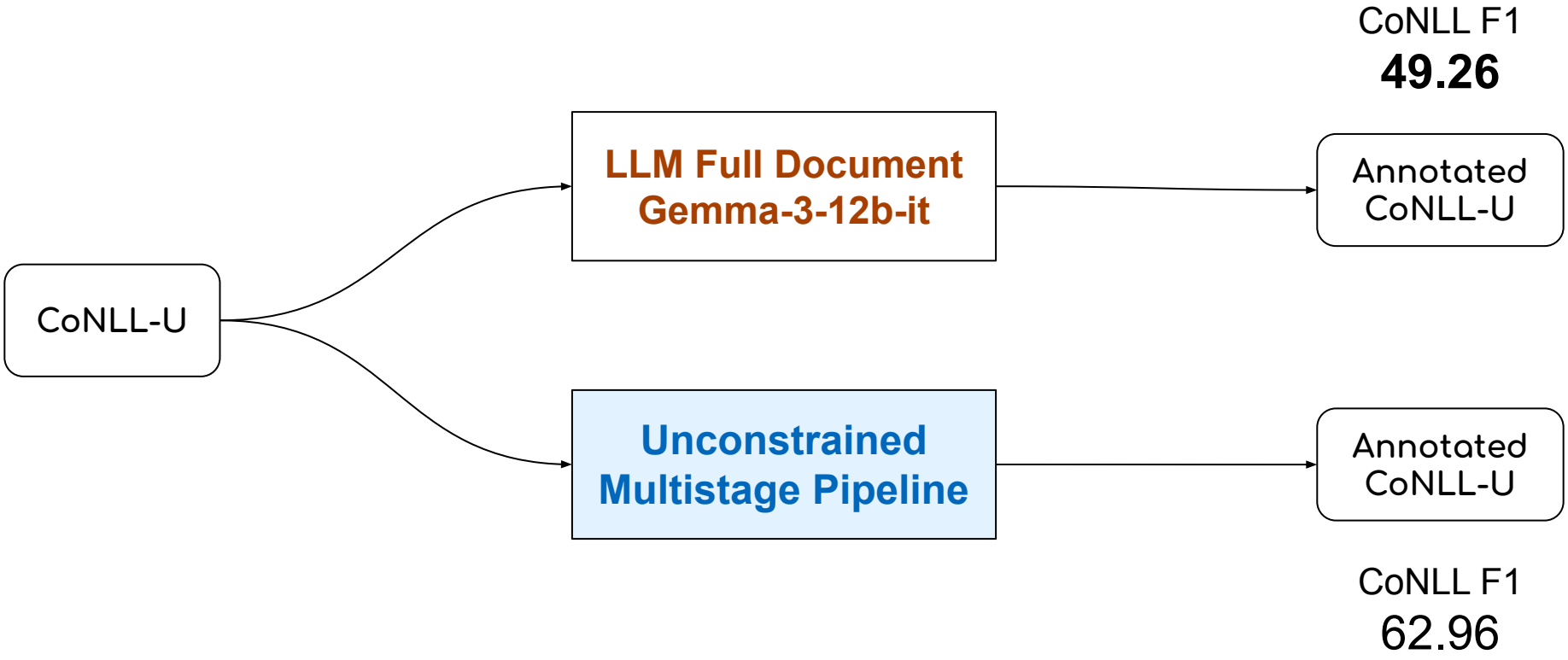
Full Document Annotation: Model Size Impact



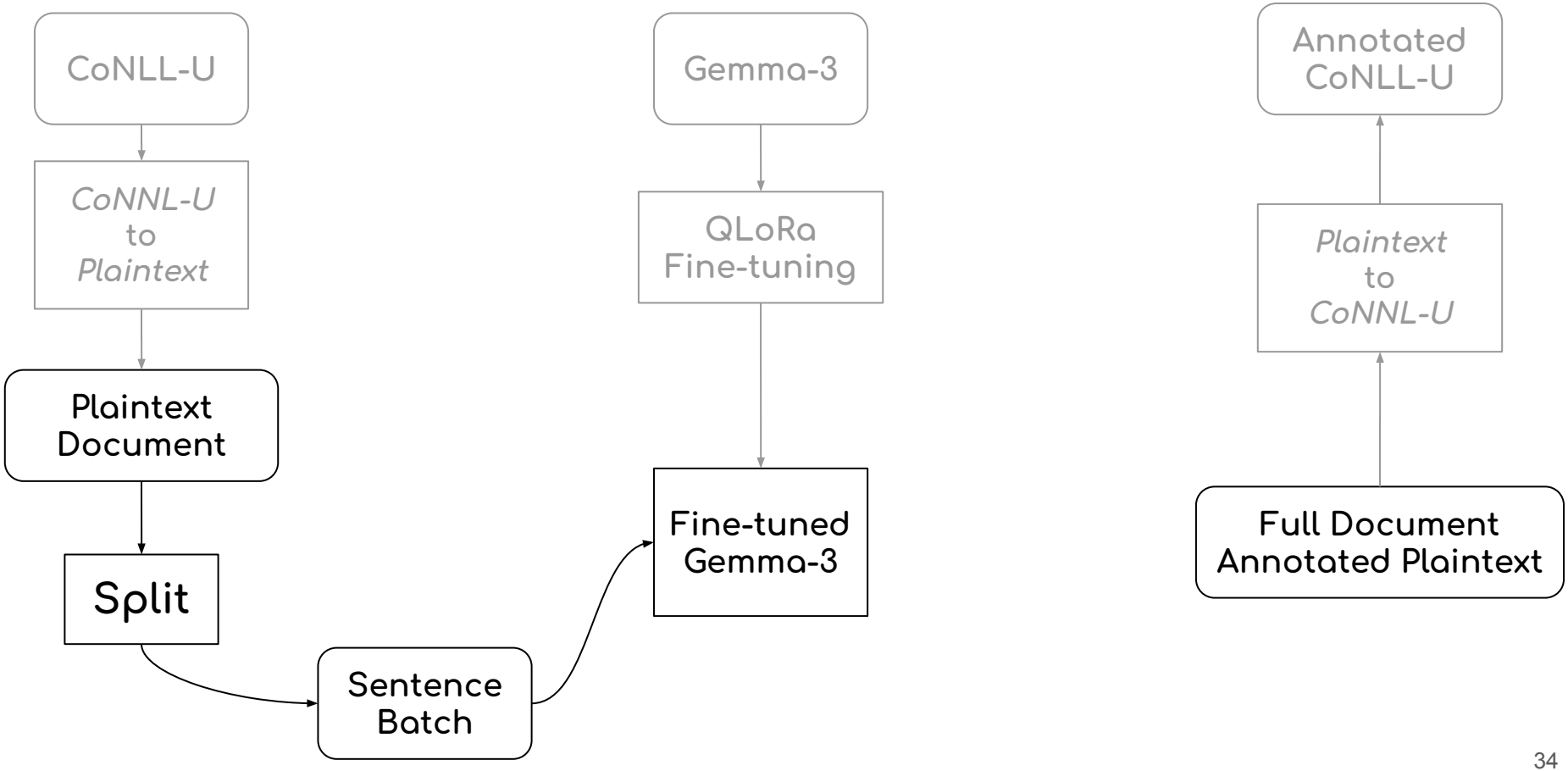
Full Document Annotation: Model Size Impact



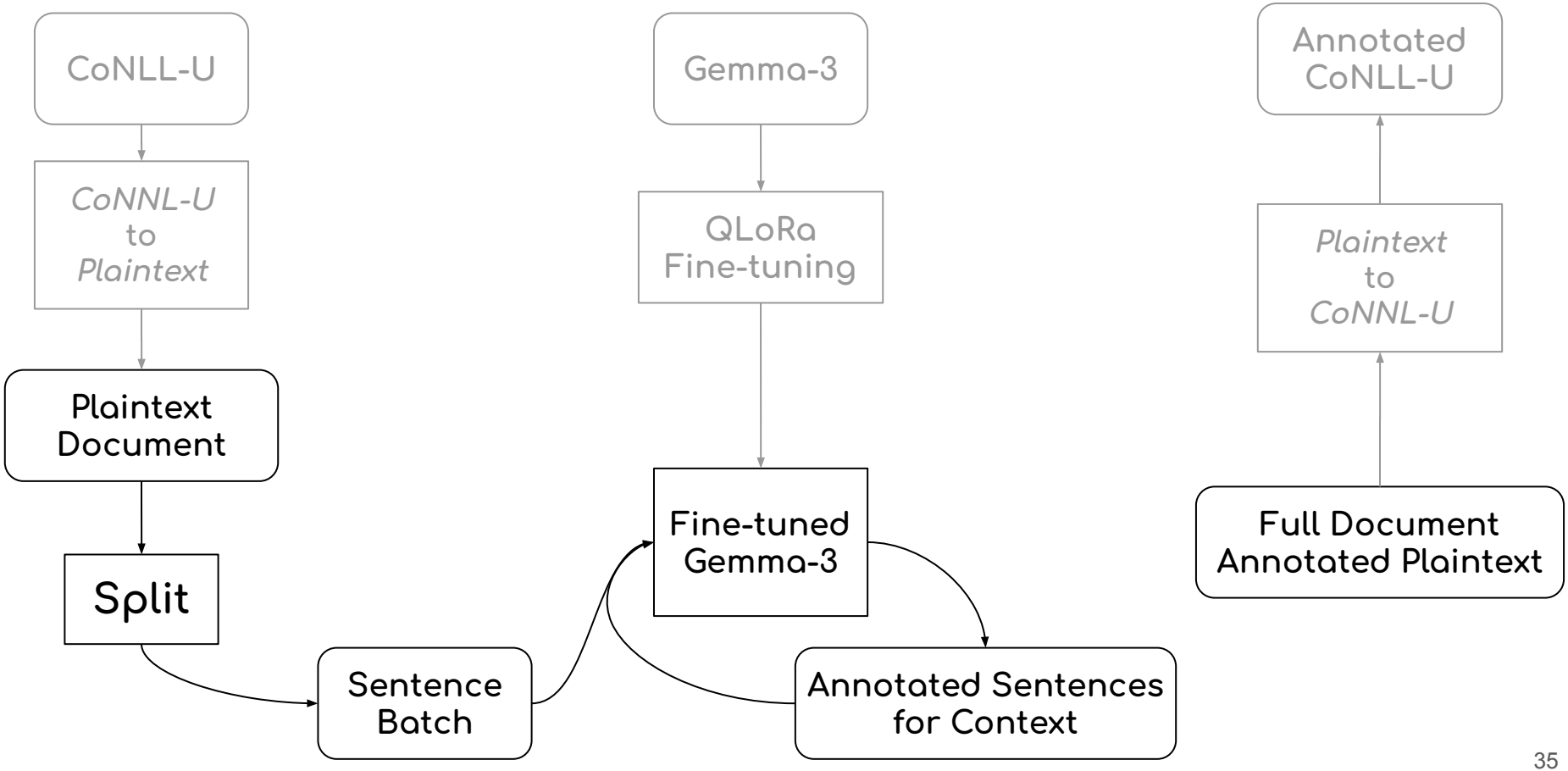
Unconstrained > LLM Full Document



Incremental Sentence Batch Annotation



Incremental Sentence Batch Annotation



Incremental Sentence Batch Annotation: Prompt Template

SYSTEM INSTRUCTION PREVIOUS CONTEXT TEXT INPUT EXPECTED MODEL OUTPUT

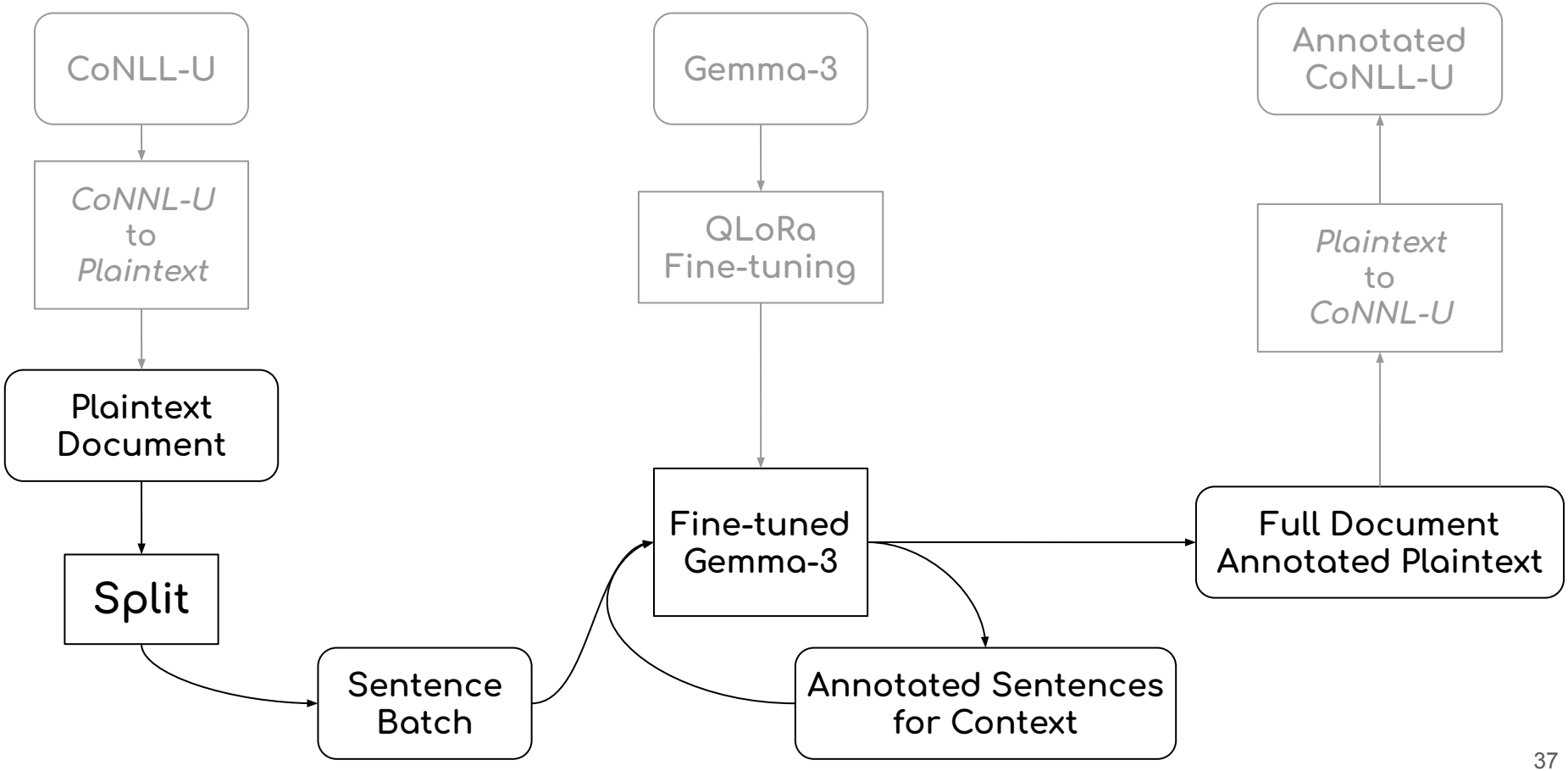
<start_of_turn>user
You are a linguist, expert in anaphora and coreference resolution.
Based on the previous context, annotate in the input sentences which nouns, pronouns and other expressions refer to the same entity.
Do only insert annotations. Do not insert extra linguistic material, nor punctuation markers and do not delete elements from the input texts.

Previous context: *ANNOTATED SENTENCES FROM PREVIOUS BATCHES*

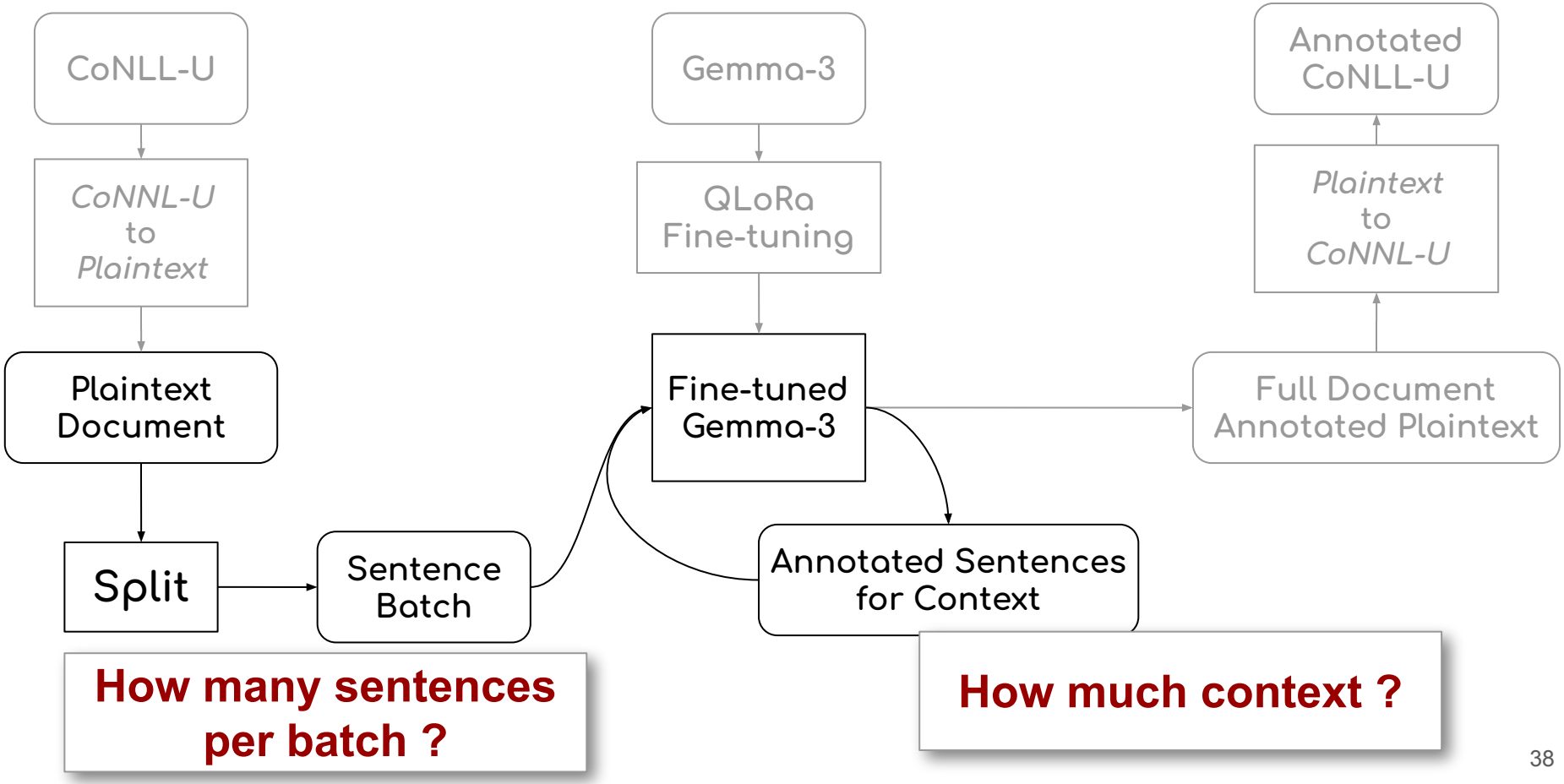
Input: *PLAINTEXT SENTENCE BATCH*
<end_of_turn>

<start_of_turn>model
COREFERENCE ANNOTATED SENTENCE BATCH
<end_of_turn>

Incremental Sentence Batch Annotation



Incremental Sentence Batch Annotation

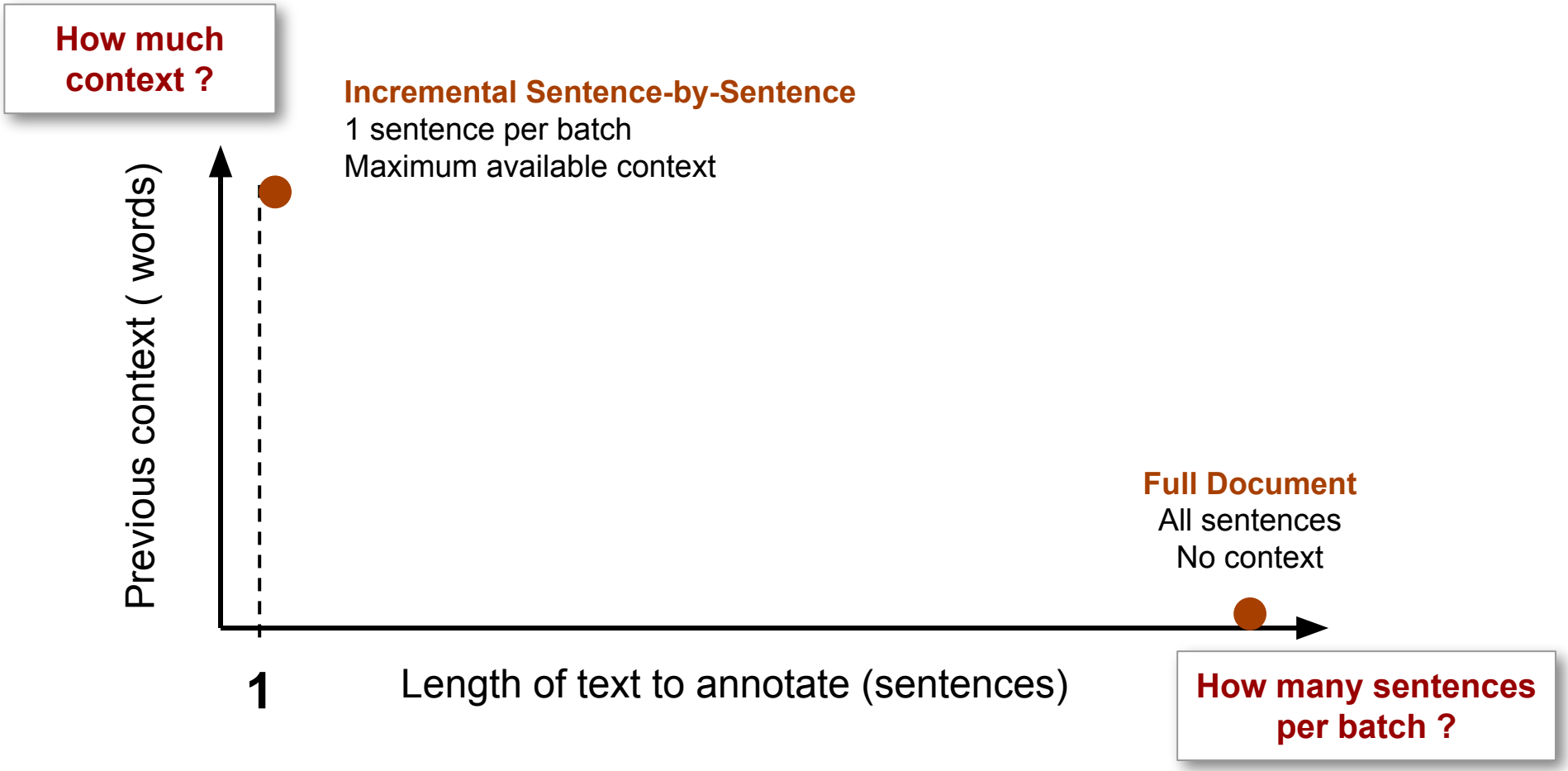


Annotation Strategies

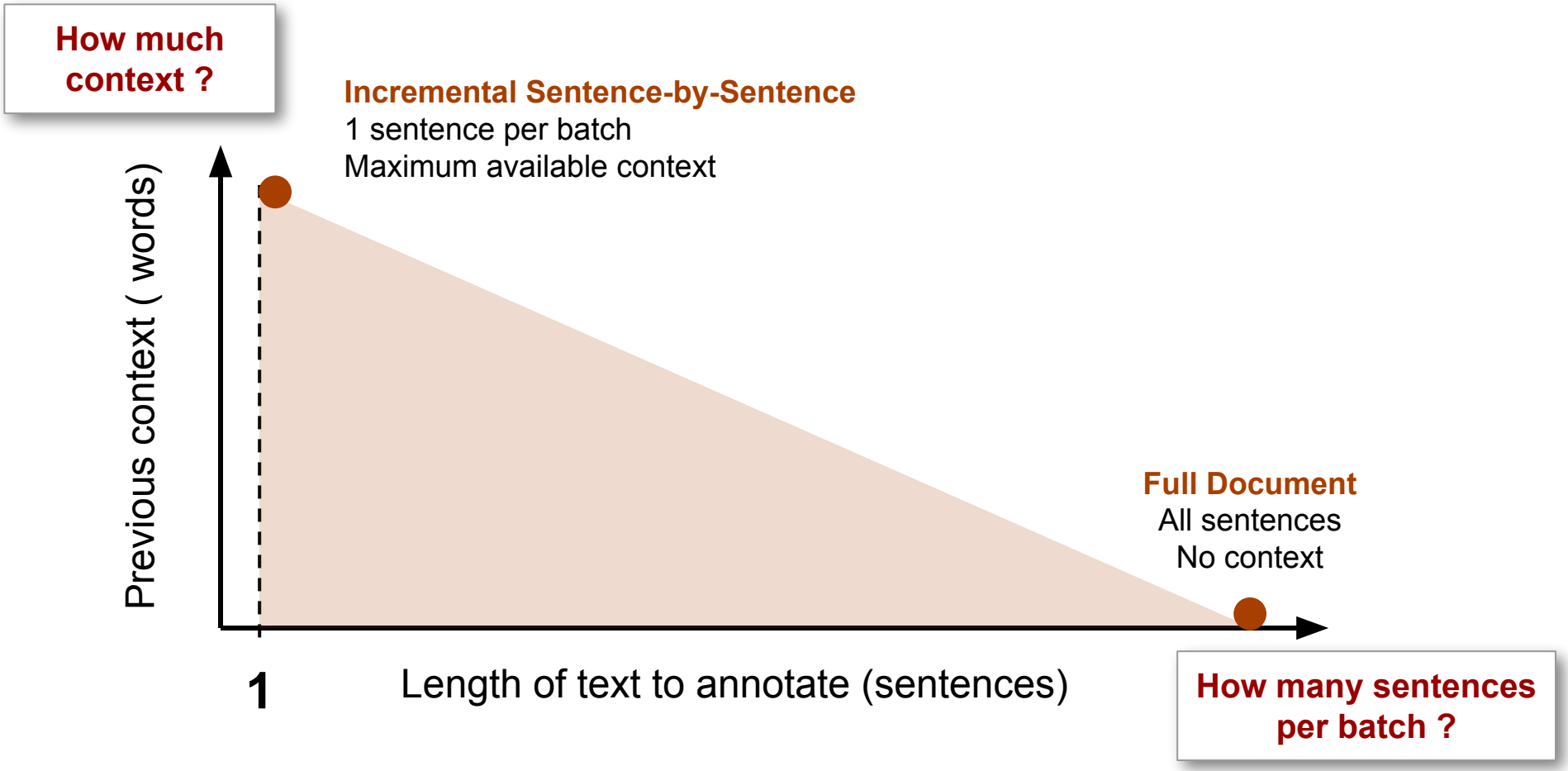
How much context ?

How many sentences per batch ?

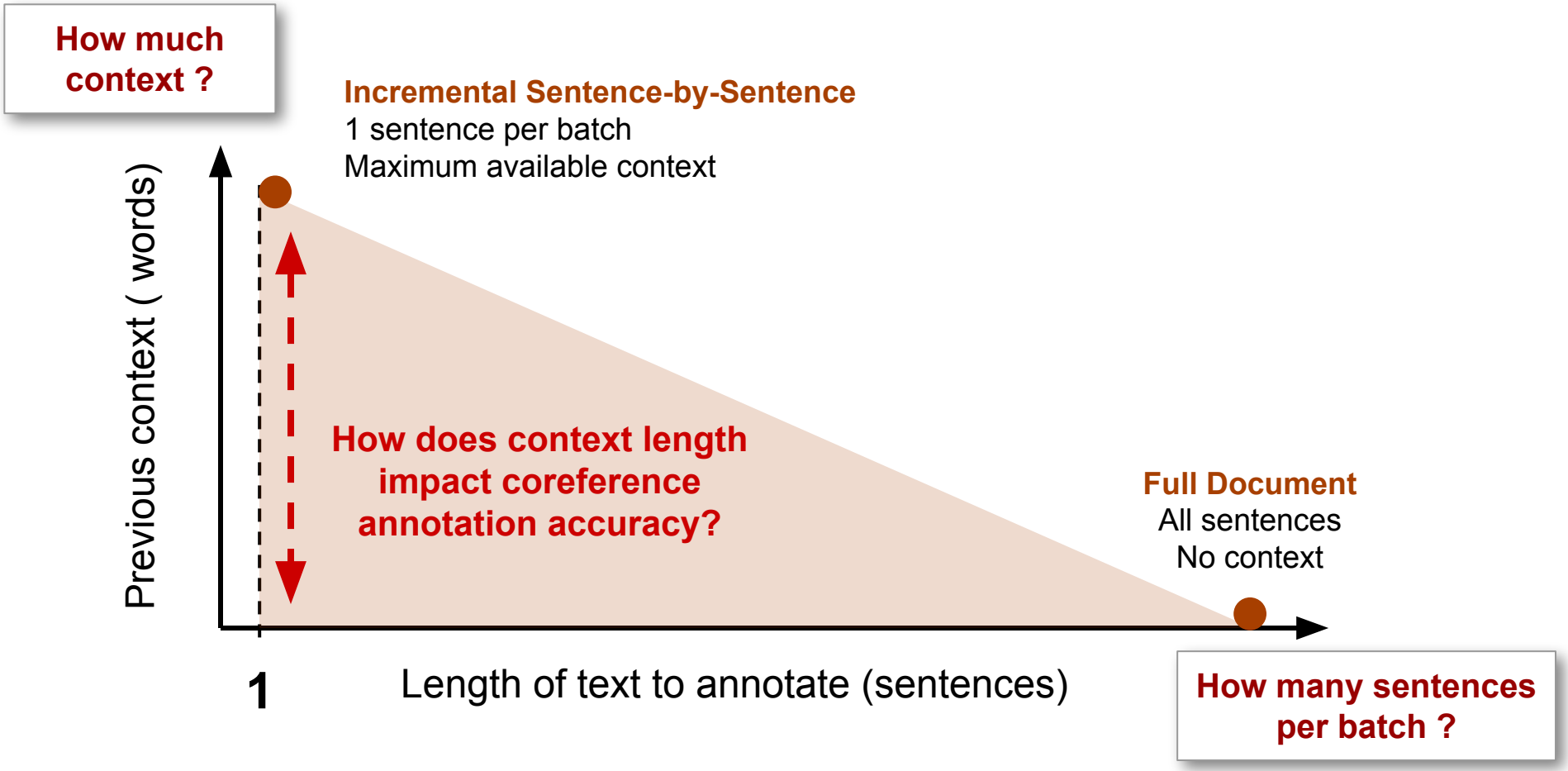
Annotation Strategies



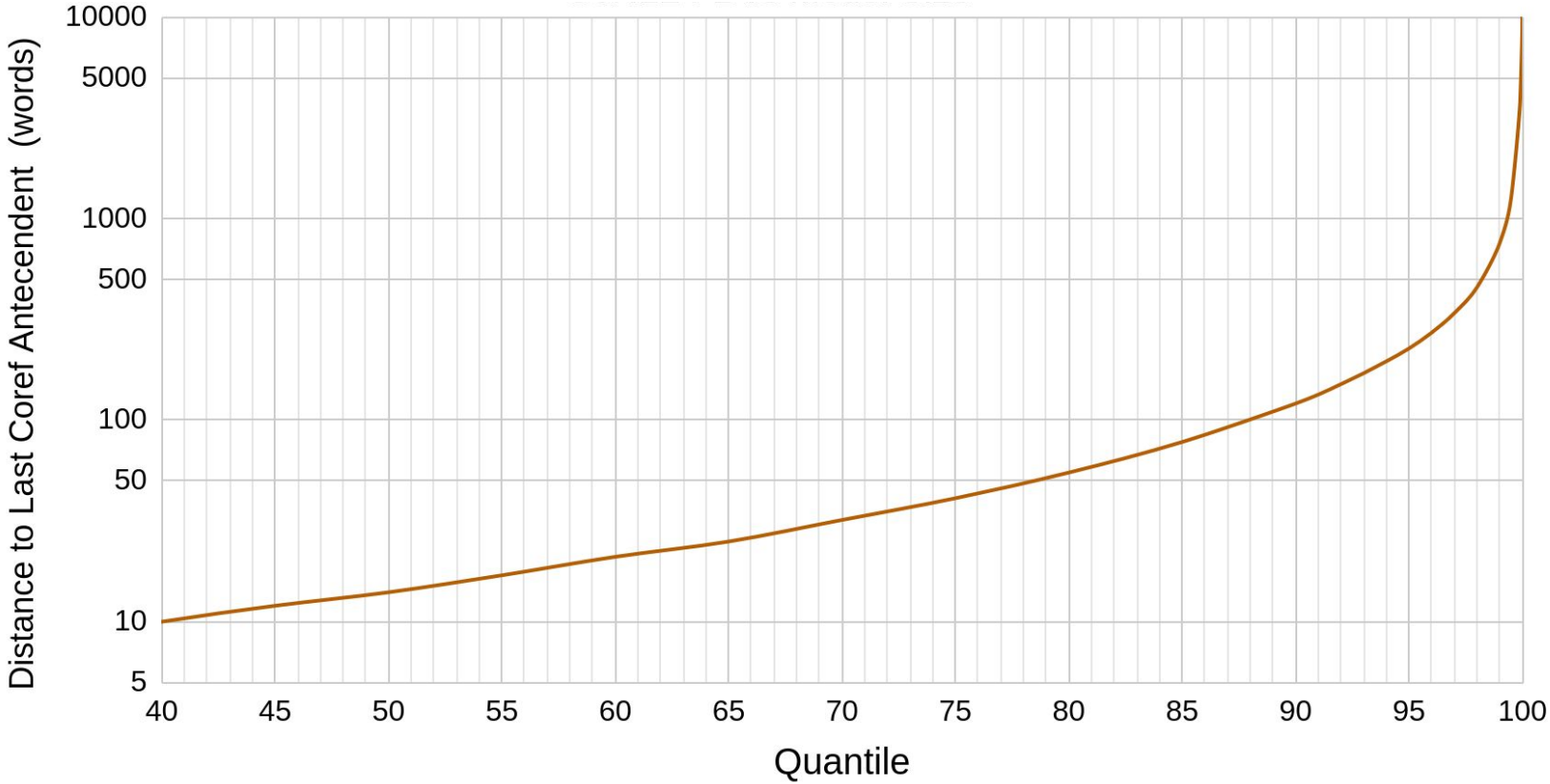
Space of Annotation Strategies



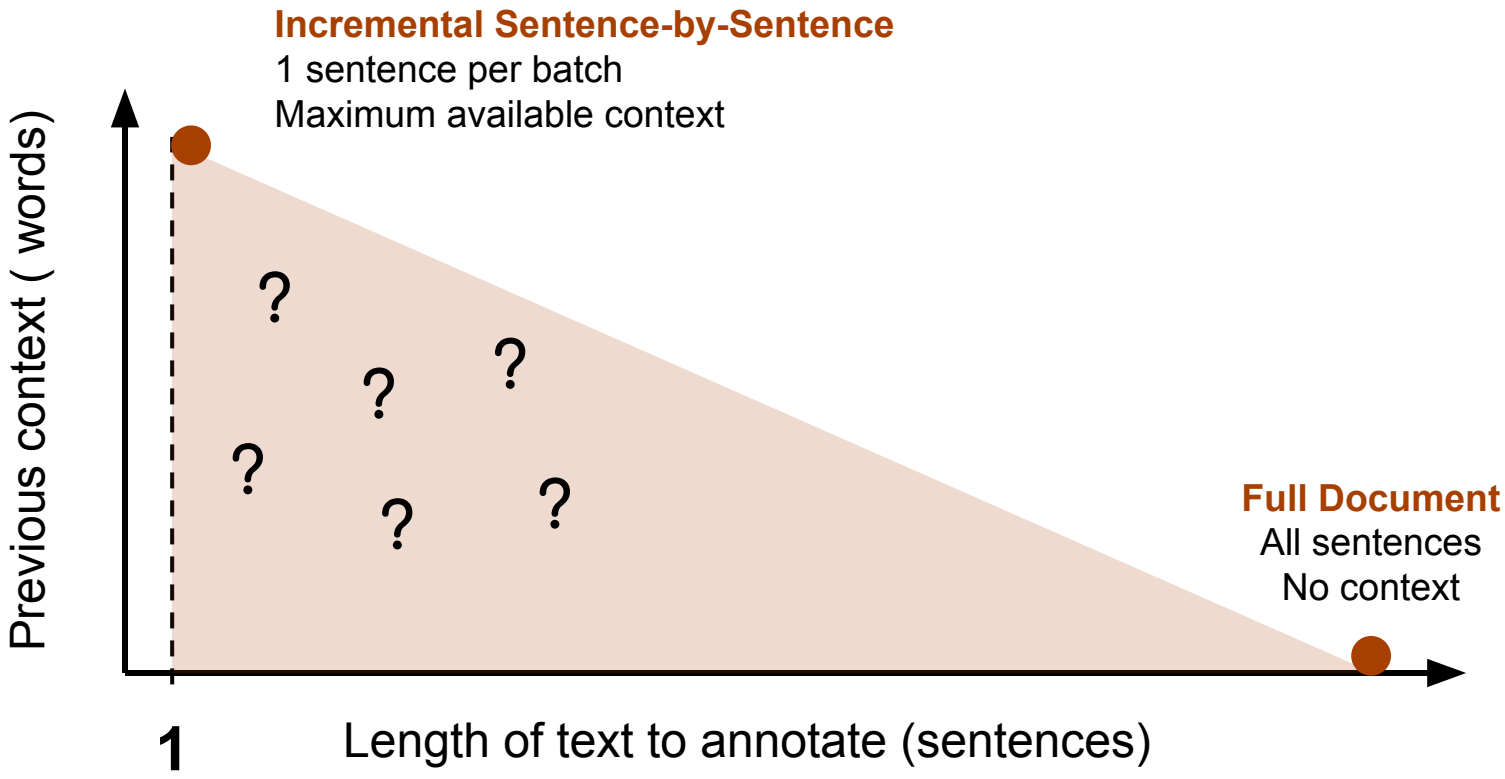
Space of Annotation Strategies



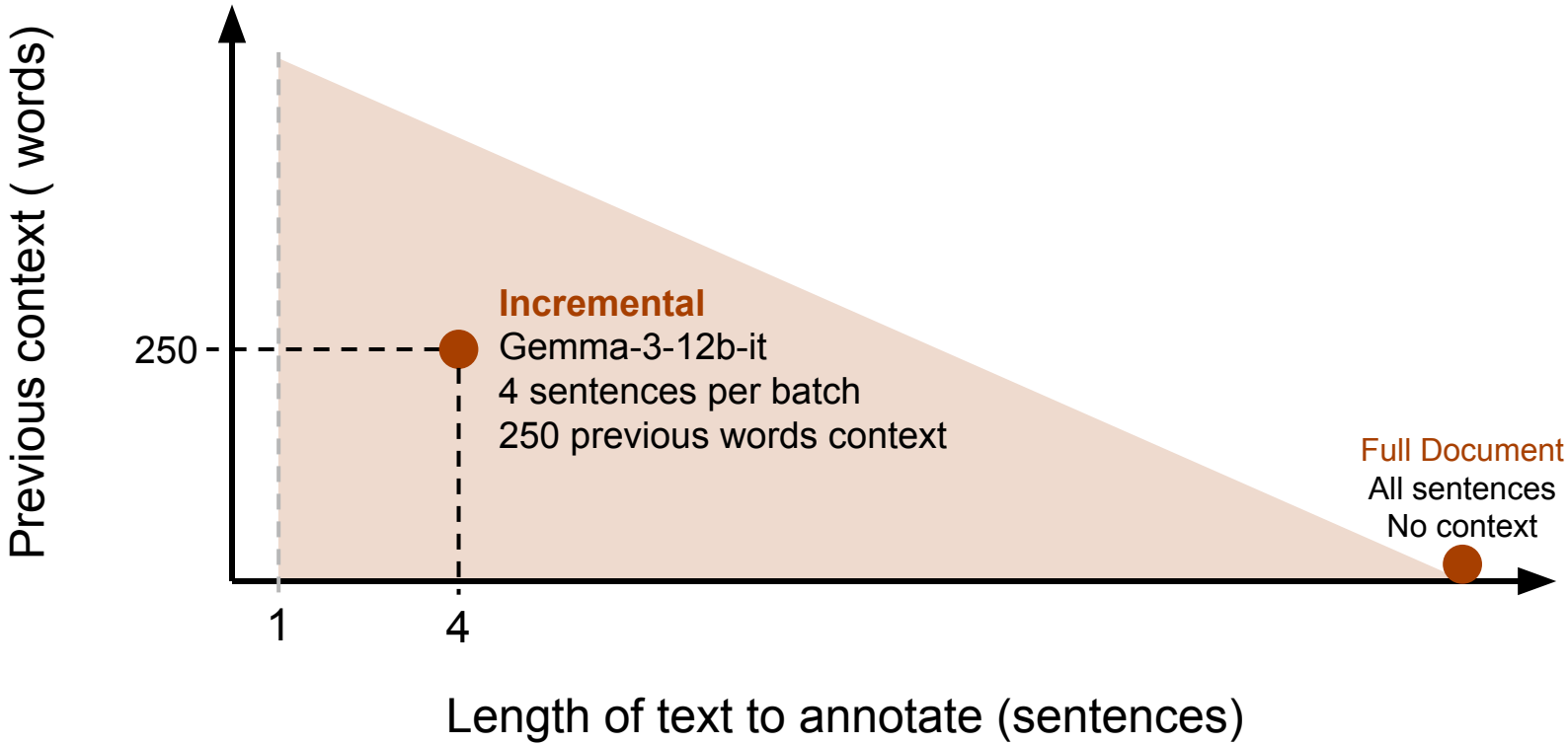
Distance to Last Coreferential Antecedent (train set)



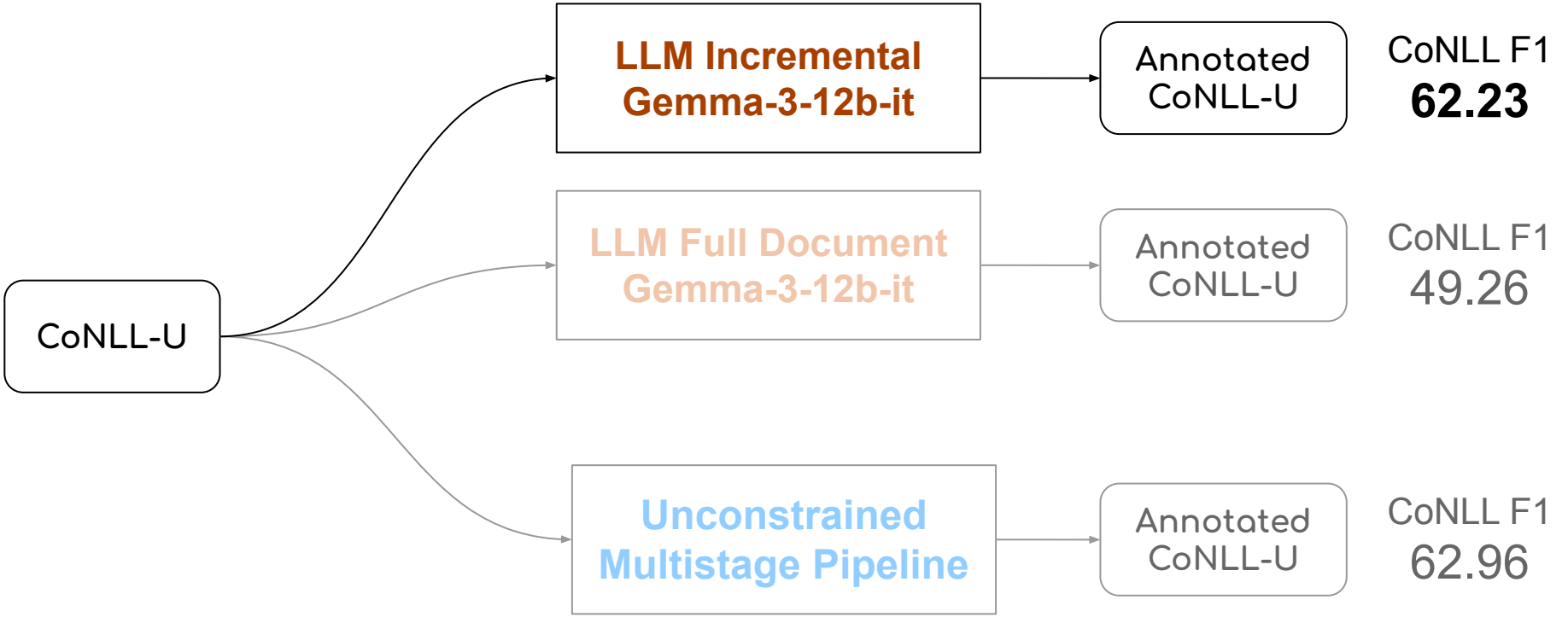
Space of Annotation Strategies: Exploratory Experiments



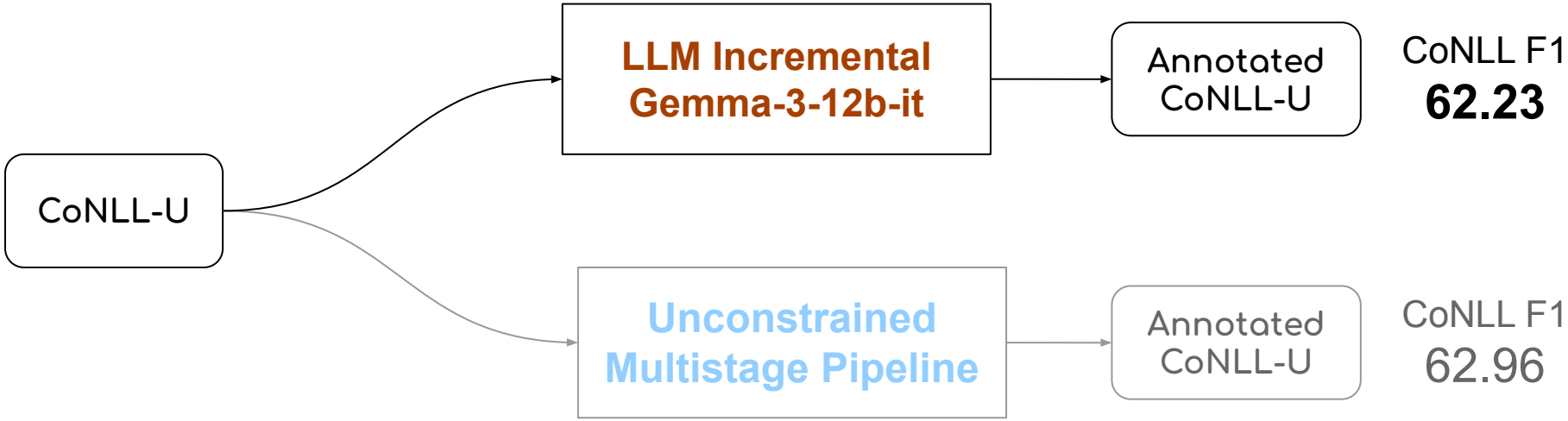
Space of Annotation Strategies



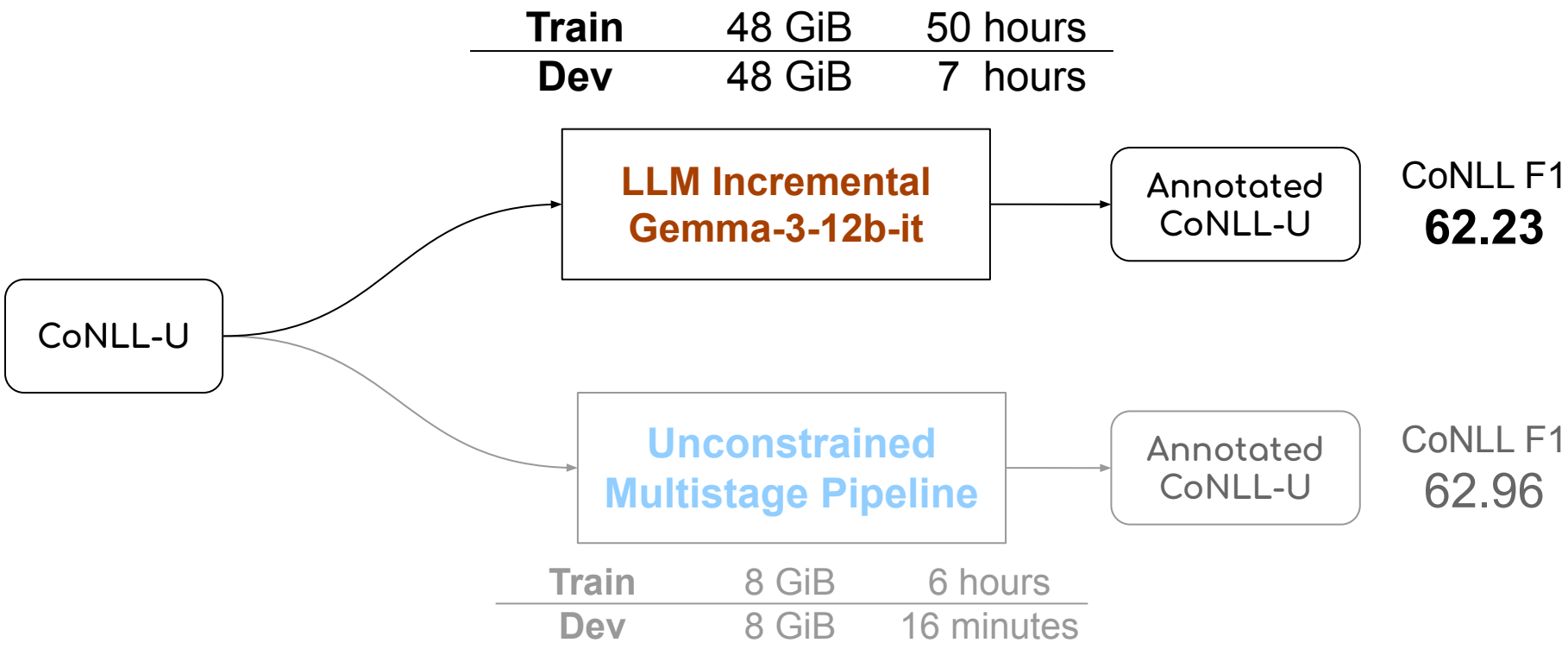
LLM Incremental: Results



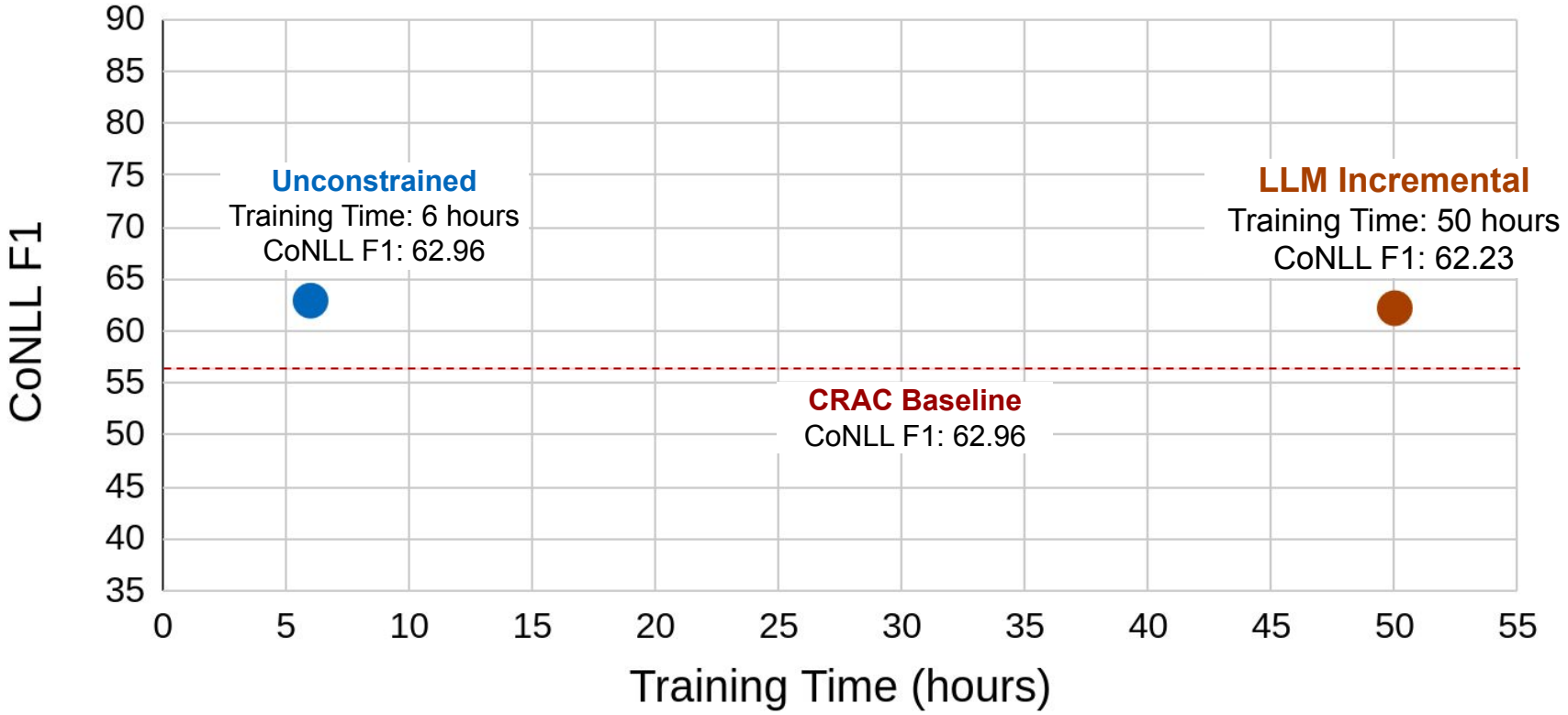
LLM Incremental: Best Model



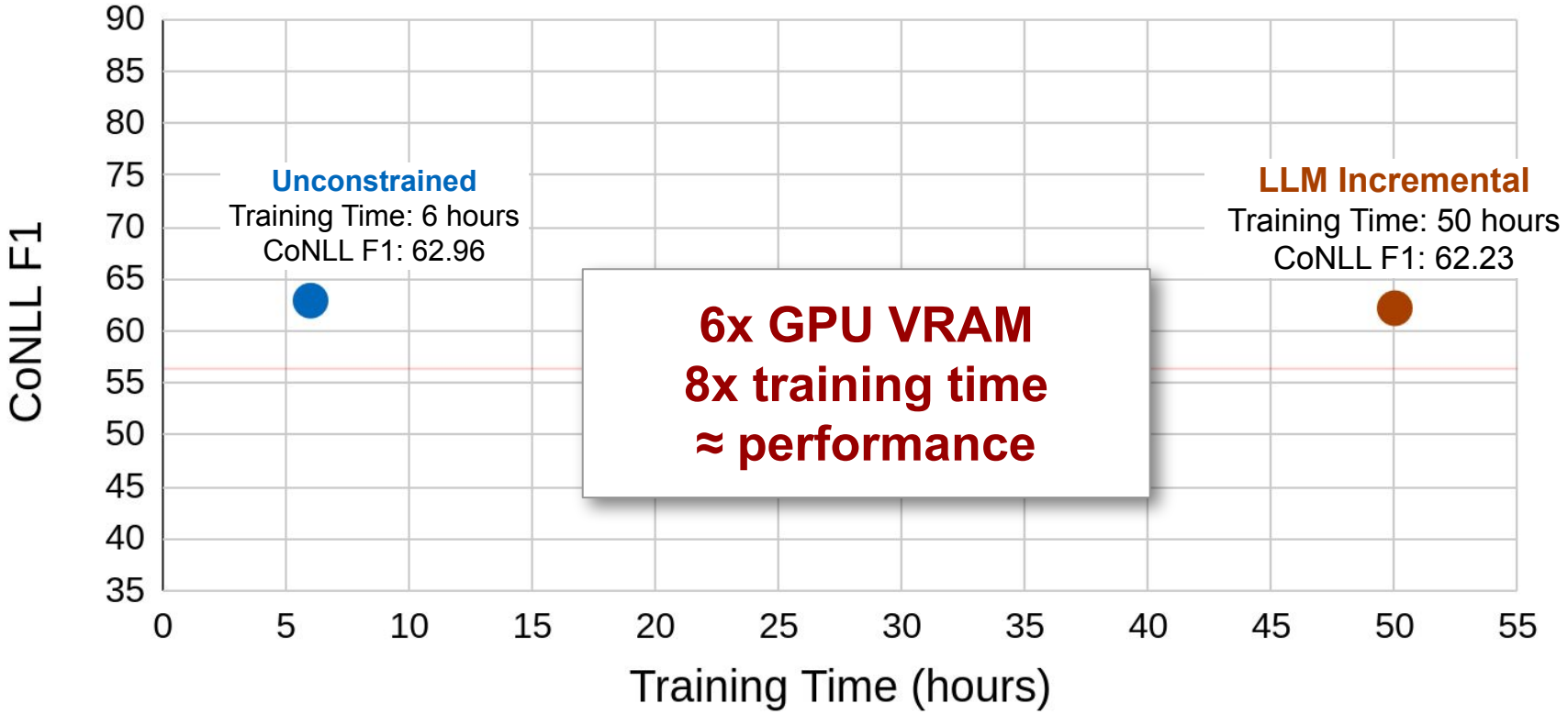
LLM Incremental: Ressources



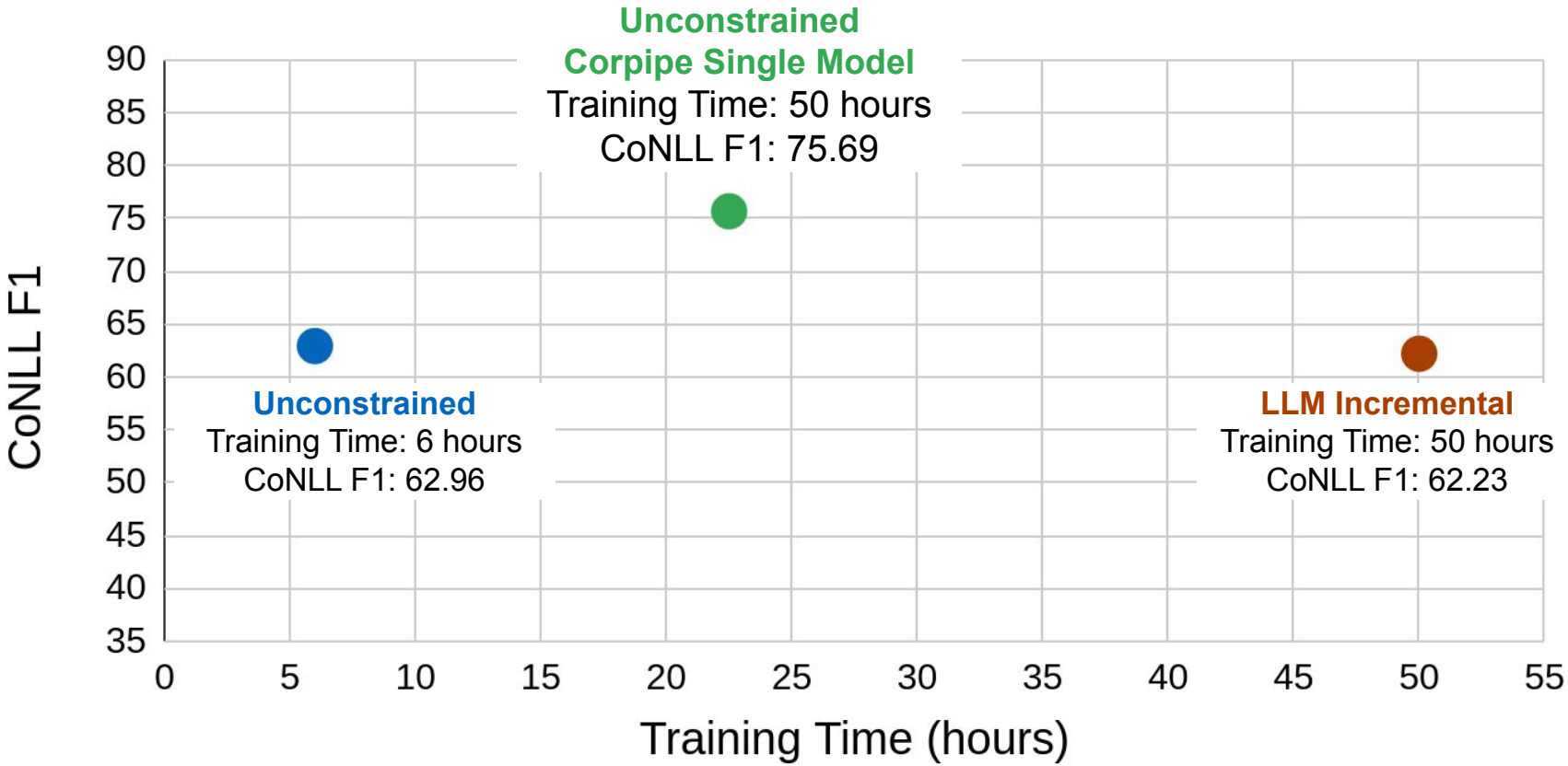
Comparison of Models



Comparison of Models



Comparison of Models: Other Participants



Comparison of Models: Other Participants

system	head-match
LLM-GLaRef-CRAC25	**62.96**
LLM-NUST-FewShot	61.74
LLM-PUXCRAC2025	60.09
LLM-UWB	59.84
CorPipeEnsemble	**75.84**
CorPipeBestDev	75.06
CorPipeSingle	74.75
Stanza	67.81
GLaRef-Propp	61.57
BASELINE-GZ	58.18
BASELINE	56.01

Findings of the Fourth Shared Task on Multilingual Coreference Resolution: Can LLMs Dethrone Traditional Approaches?

Comparison of Models: Other Participants

system	head-match
LLM-GLaRef-CRAC25	**62.96**
LLM-NUST-FewShot	61.74
LLM-PUXCRAC2025	60.09
LLM-UWB	59.84
CorPipeEnsemble	**75.84**
CorPipeBestDev	75.06
CorPipeSingle	74.75
Stanza	67.81
GLaRef-Propp	61.57
BASELINE-GZ	58.18
BASELINE	56.01

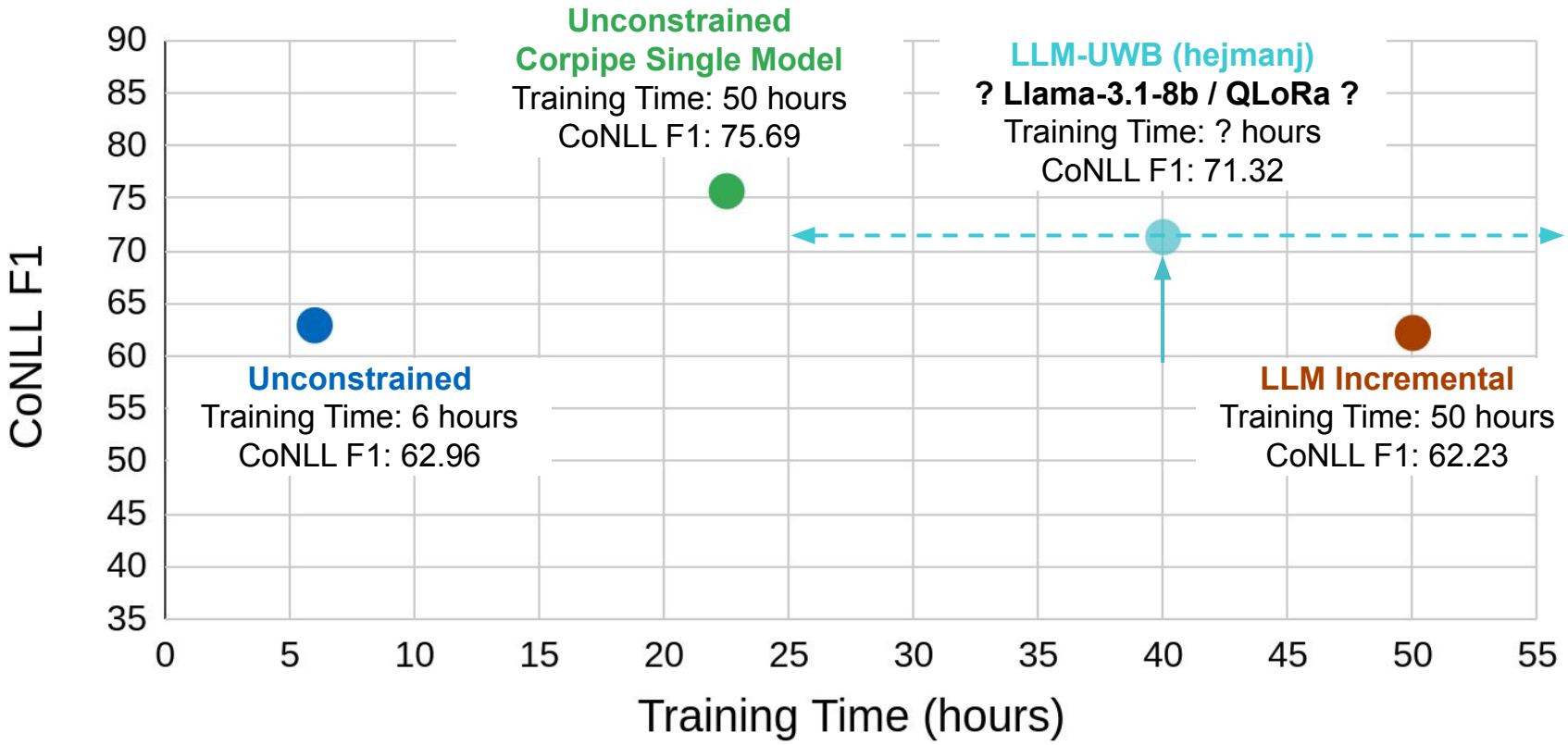
Findings of the Fourth Shared Task on Multilingual Coreference Resolution: Can LLMs Dethrone Traditional Approaches?

LLM-UWB (hejmanj) The UWB team fine-tunes a Llama-3.1-8B model on the official plain-text export of the CoNLL-U files. Training is done using QLoRA adaptation. The model is trained to

#	User	Entries	Date of Last Entry	avg ▲
1	hejmanj	5	07/29/25	71.32 (1)

Submitted after the official deadline

Comparison of Models: Other Participants

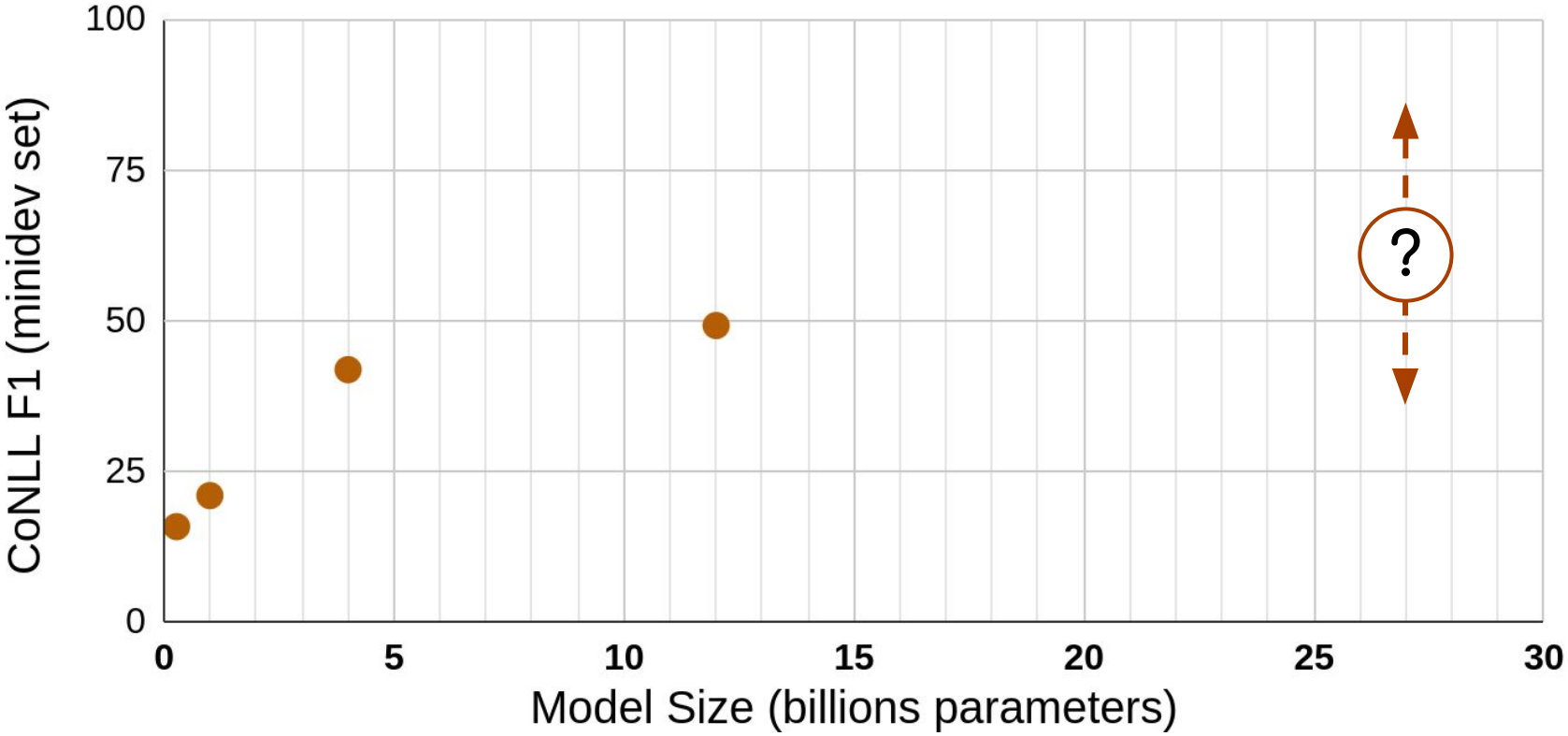


LLM Coreference Resolution: Perspectives

LLM Coreference Resolution: Perspectives

1. Model Size Increase
2. Coreference-aware Loss Function
3. Coreference ID Tracking
4. Plaintext Format

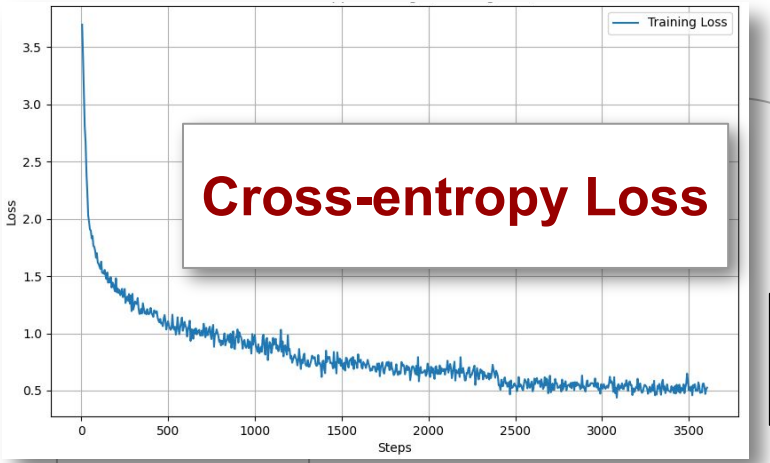
Perspectives: Model Size Increase



LLM Coreference Resolution: Perspectives

1. Model Size Increase
2. Coreference-aware Loss Function
3. Coreference ID Tracking
4. Plaintext Format

Perspectives: Coreference-aware Loss Function



Annotated CoNLL-U

Are incorrect coreference IDs properly penalized?

QLoRa Fine-tuning

Plaintext to CoNNL-U

So she|[e2] was considering in her|[e2] own mind (as well as she|[e2] could , for the hot day made her|[e2] feel very sleepy and stupid)

Plaintext

Fine-tuned

Annotated

So she|[e2] was considering in her|[e2] own mind (as well as she|[e3] could , for the hot day made her|[e2] feel very sleepy and stupid)

LLM Coreference Resolution: Perspectives

1. Model Size Increase
2. Coreference-aware Loss Function
3. Coreference ID Tracking
4. Plaintext Format

Perspectives: Coreference ID Tracking

SYSTEM INSTRUCTION PREVIOUS CONTEXT TEXT INPUT EXPECTED MODEL OUTPUT

```
<start_of_turn>user
You are a linguist, expert in anaphora and coreference resolution.
Based on the previous context, annotate in the input sentences which nouns,
pronouns and other expressions refer to the same entity.
Do only insert annotations. Do not insert extra linguistic material, nor
punctuation markers and do not delete elements from the input texts.

Previous context: *ANNOTATED SENTENCES FROM PREVIOUS BATCHES*

Maximum ID used: [e13]

Input: *PLAINTEXT SENTENCE BATCH*
<end_of_turn>

<start_of_turn>model
*COREFERENCE ANNOTATED SENTENCE BATCH*
<end_of_turn>
```

Perspectives: Coreference ID Tracking

SYSTEM INSTRUCTION PREVIOUS CONTEXT TEXT INPUT EXPECTED MODEL OUTPUT

```
<start_of_turn>user
You are a linguist, expert in anaphora and coreference resolution.
Based on the previous context, annotate in the input sentences which nouns,
pronouns and other expressions refer to the same entity.
Do only insert annotations. Do not insert extra linguistic material, nor
punctuation markers and do not delete elements from the input texts.

Previous context: *ANNOTATED SENTENCES FROM PREVIOUS BATCHES*

Entity Tracker: "Alice's sister"[e1], "Alice"[e2], "The White Rabbit"[e3]

Input: *PLAINTEXT SENTENCE BATCH*
<end_of_turn>

<start_of_turn>model
*COREFERENCE ANNOTATED SENTENCE BATCH*
<end_of_turn>
```

LLM Coreference Resolution: Perspectives

1. Model Size Increase
2. Coreference-aware Loss Function
3. Coreference ID Tracking
4. Plaintext Format

Perspectives: Plaintext format conversion

Down the**|[e1** Rabbit-Hole**|e1]** Alice**|[e2]** was beginning to get very tired of sitting by her**|[e2],[e3** sister**|e3]** on the **|[e4** bank**|e4]** , and of having nothing to do : once or twice she**|[e2]** had peeped into the book her**|[e2],[e3** sister**|e3]** was reading.

Down **<e1>**the Rabbit-Hole**</e1>** **<e2>**Alice**</e2>** was beginning to get very tired of sitting by **<e3><e2>**her**</e2>** sister**</e3>** on **<e4>**the bank**</e4>** , and of having nothing to do : once or twice **<e2>**she**</e2>** had peeped into thebook **<e3><e2>**her**</e2>** sister**</e3>** was reading.

Alternative tagging scheme inspired by markup languages like HTML or XML that tokenizers and LLMs might be more familiar with.

Perspectives: Coreference ID Tracking

1. Model Size Increase
2. Coreference-aware Loss Function
3. Coreference ID Tracking
4. Plaintext Format
5. **Other Suggestions ?**

Thank you for your attention

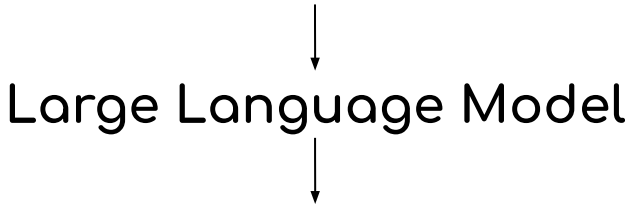
Antoine BOURGOIS
Lattice (CNRS – École Normale Supérieure – Université Sorbonne Nouvelle), Paris, France
antoine.bourgois@ens.psl.eu

Additional Material: Incremental approach

PREVIOUS CONTEXT RAW TEXT INPUT EXPECTED MODEL OUTPUT

STEP 1

[None] CHAPTER I. Down the Rabbit-Hole Alice was beginning to get very tired of sitting by her sister on the bank , and of having nothing to do : once or twice she had peeped into the book her sister was reading , but it had no pictures or conversations in it , ‘ and what is the use of a book , ’ thought Alice ‘ without pictures or conversations ? ’



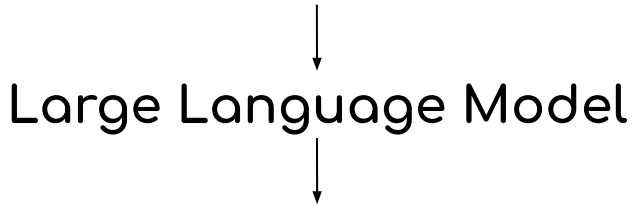
CHAPTER I. Down the|[e1 Rabbit-Hole|e1] Alice|[e2] was beginning to get very tired of sitting by her|[e2],[e3 sister|e3] on the|[e4 bank|e4] , and of having nothing to do : once or twice she|[e2] had peeped into the book her|[e2],[e3 sister|e3] was reading , but it had no pictures or conversations in it , ‘ and what is the use of a book , ’ thought Alice|[e2] ‘ without pictures or conversations ? ’

Additional Material: Incremental approach

PREVIOUS CONTEXT RAW TEXT INPUT EXPECTED MODEL OUTPUT

STEP 2

CHAPTER I. Down the|[e1 Rabbit-Hole|e1] Alice|[e2] was beginning to get very tired of sitting by her|[e2],[e3 sister|e3] on the|[e4 bank|e4] , and of having nothing to do : once or twice she|[e2] had peeped into the book her|[e2],[e3 sister|e3] was reading , but it had no pictures or conversations in it , ‘ and what is the use of a book , ’ thought Alice|[e2] ‘ without pictures or conversations ? ’ So she was considering in her own mind (as well as she could , for the hot day made her feel very sleepy and stupid) , whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies , when suddenly a White Rabbit with pink eyes ran close by her .



So she|[e2] was considering in her|[e2] own mind (as well as she|[e2] could , for the hot day made her|[e2] feel very sleepy and stupid) , whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies , when suddenly a|[e5 White Rabbit with pink eyes|e5] ran close by her|[e2] .