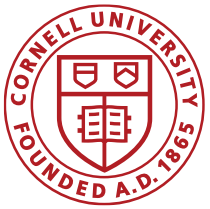
{ Graph-based & Transition-based

Combining Global Models for Parsing Universal Dependencies

Team C2L2 —

Tianze Shi, Felix G. Wu, Xilun Chen, Yao Cheng



Cornell University

Overview — Scope of Our System

What we did

- Projective Parsing
- Dependency Arc Labeling
- Delexicalized Parsing

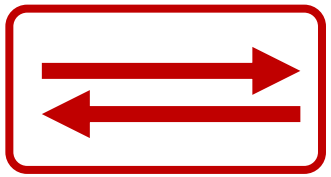
What we didn't do

- Word Segmentation
- Sentence Boundary Detection
- POS Tagging
- Morphology Analysis
- Non-projective Parsing
- Unlabeled data

Overview — Highlights

$$\underset{y \in \mathcal{Y}}{\operatorname{argmax}}$$

- Global transition-based models



- Bi-LSTM-powered compact features

2nd

- Overall

fi → sme

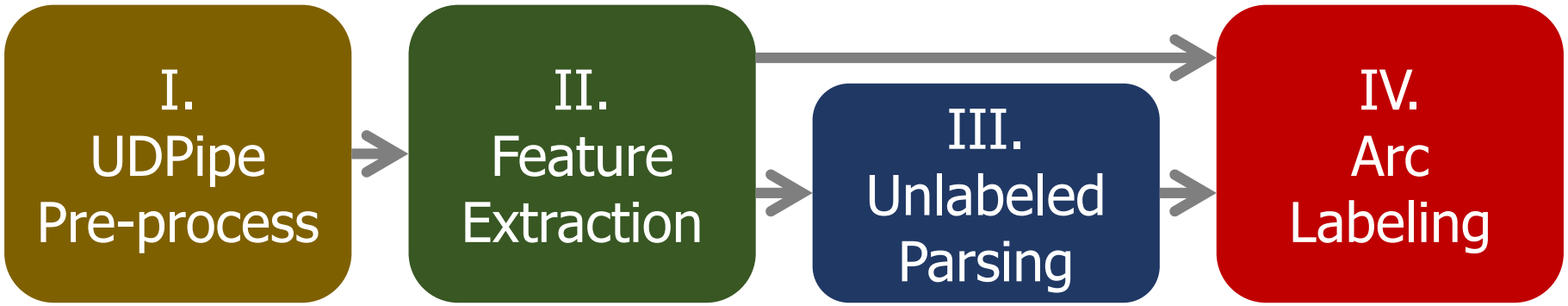
- Delexicalized syntactic transfer

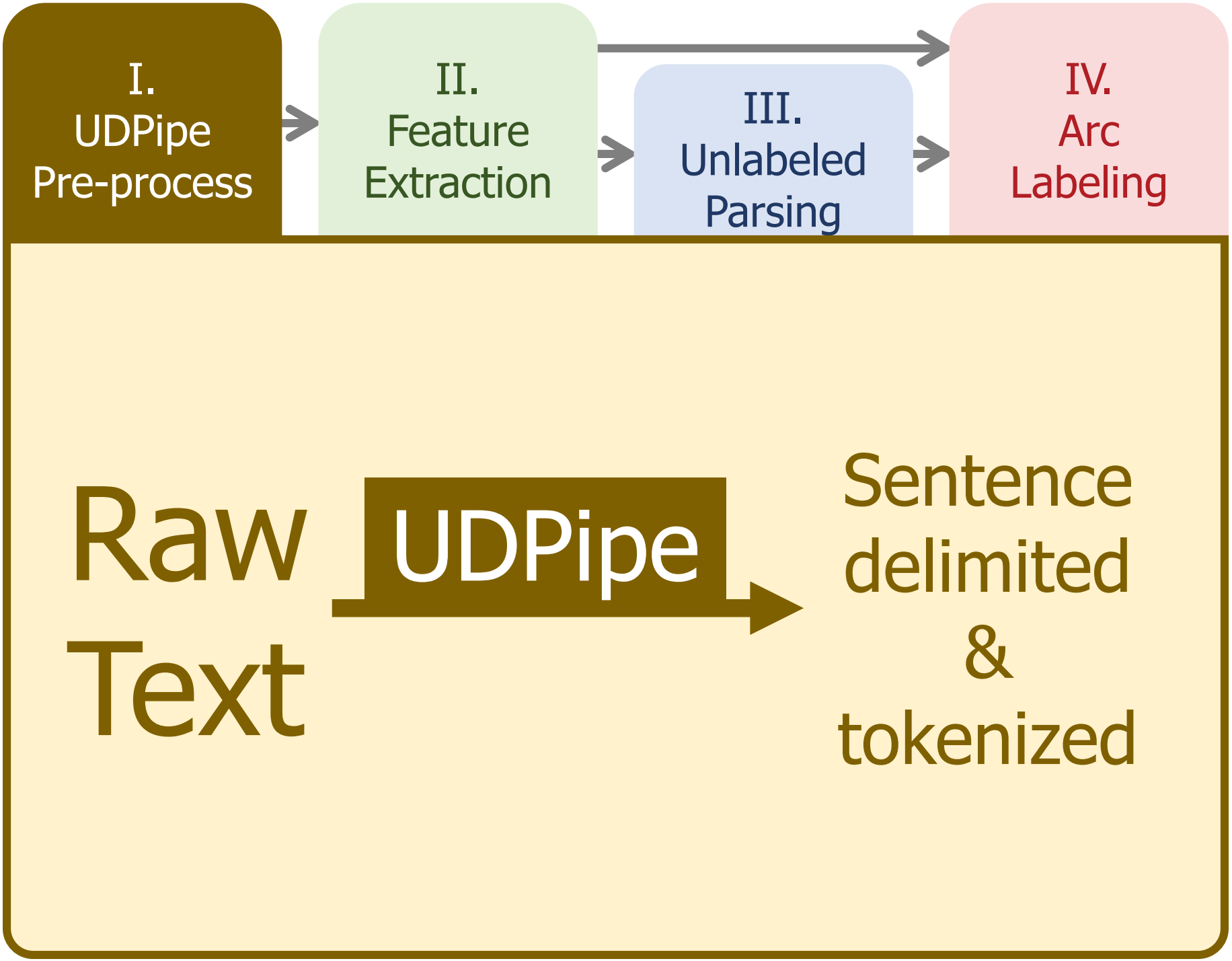
- High efficiency, low resource demand

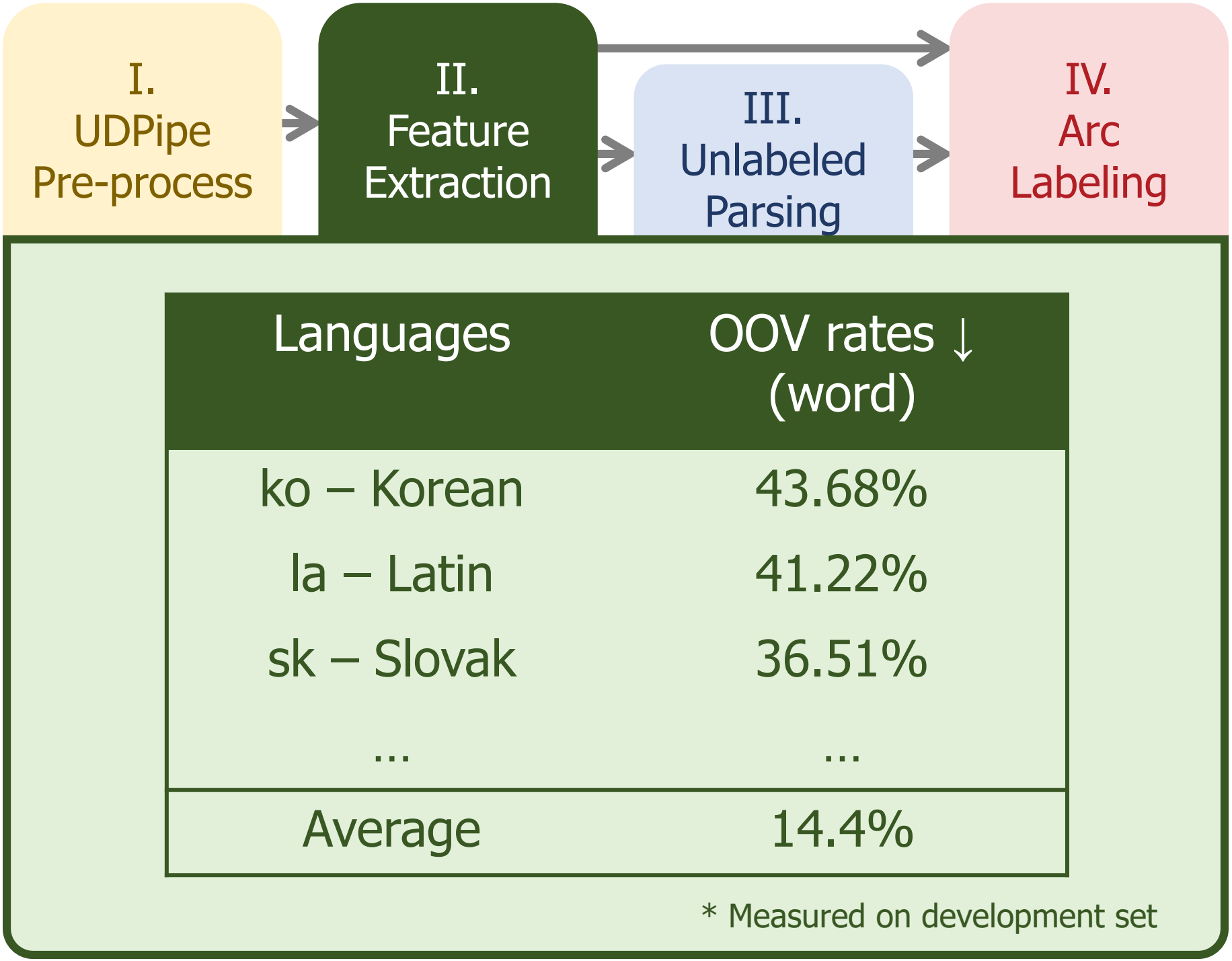
1st

- Small Treebanks
- Surprise Languages

Overview — System Pipeline

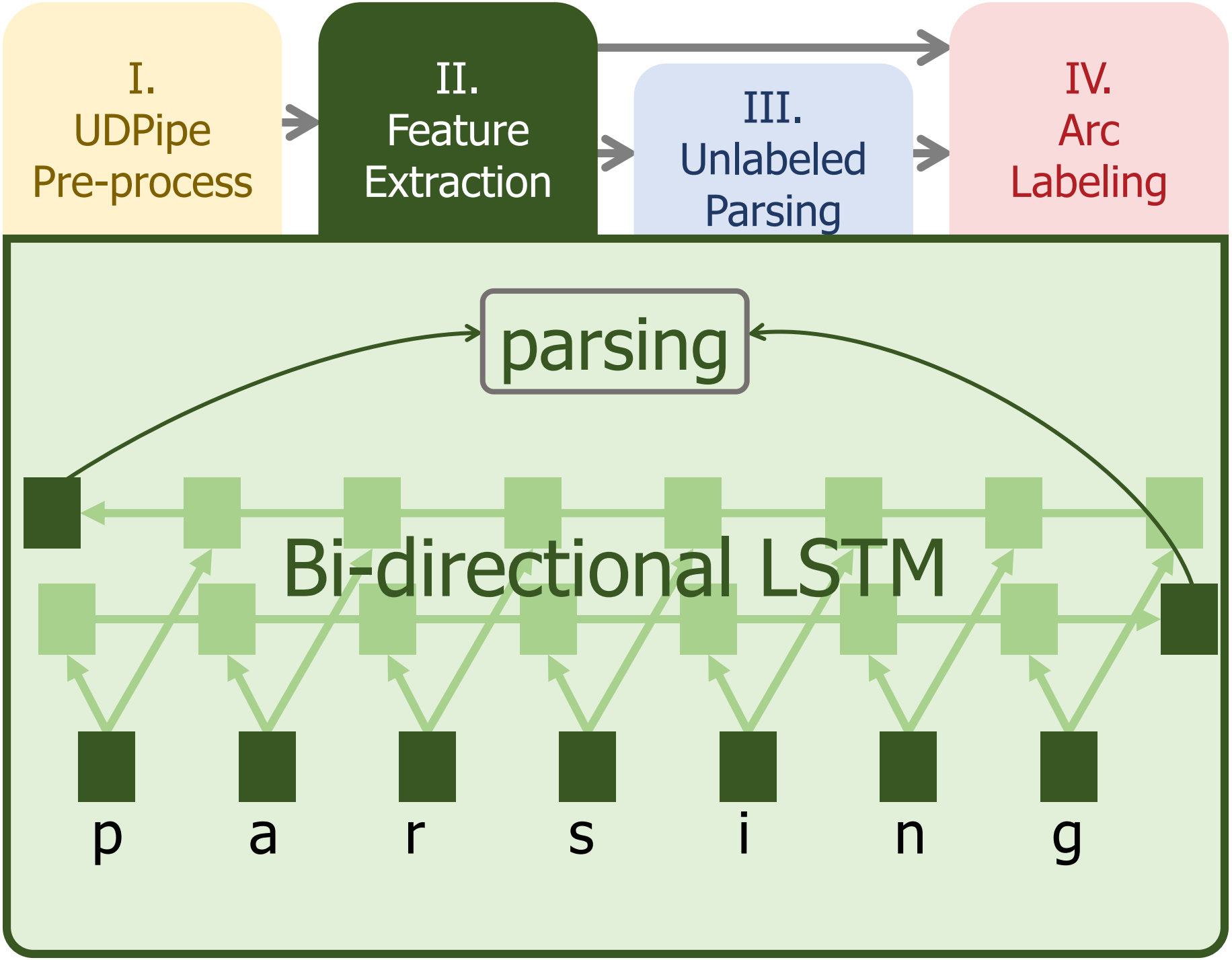


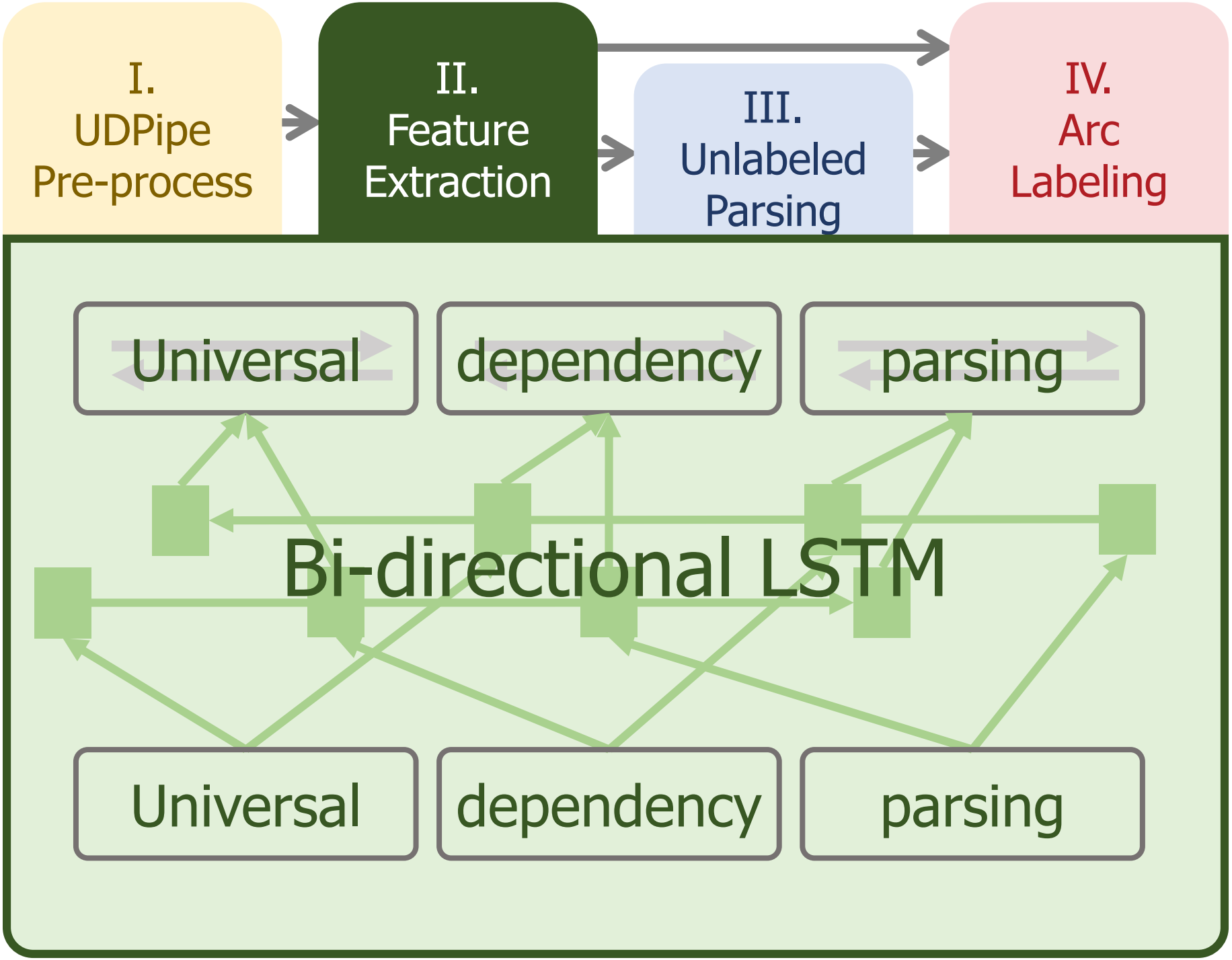

I.
UDPipe
Pre-process
II.
Feature
Extraction
III.
Unlabeled
Parsing
IV.
Arc
Labeling
Raw
Text
UDPipe
Sentence
delimited
&
tokenized

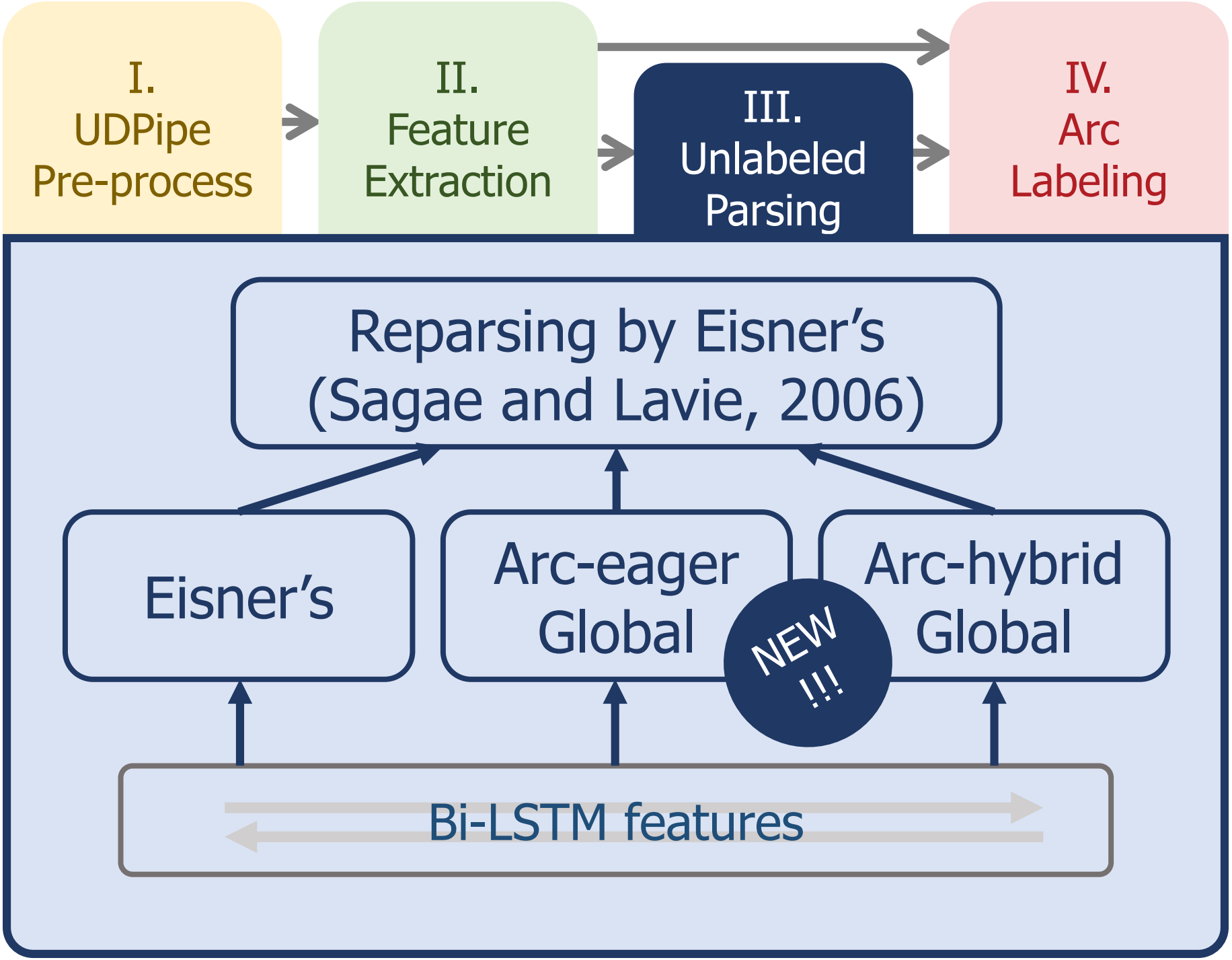


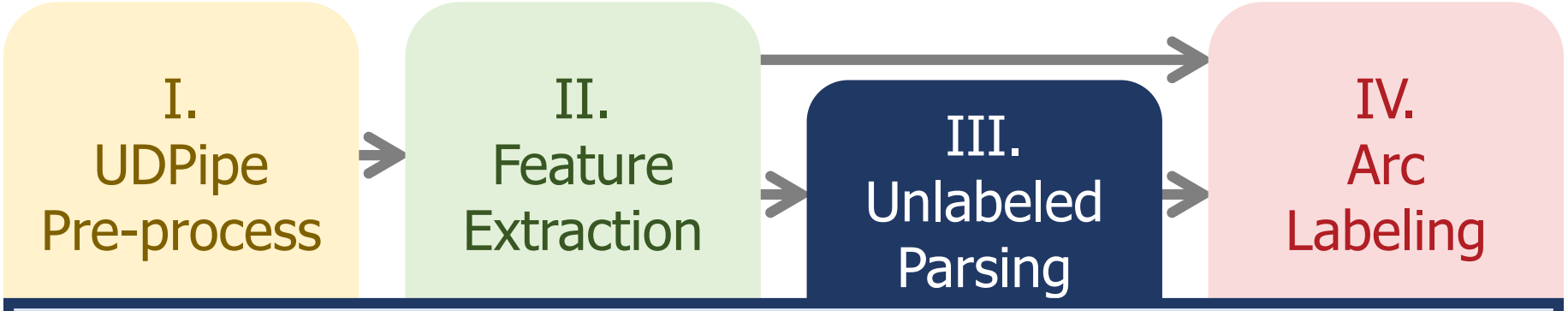
Languages	OOV rates ↓ (word)
ko – Korean	43.68%
la – Latin	41.22%
sk – Slovak	36.51%
…	…
Average	14.4%

* Measured on development set


I.
UDPipe
Pre-process
II.
Feature
Extraction
III.
Unlabeled
Parsing
IV.
Arc
Labeling
parsing
Bi-directional LSTM
p
a
r
s
i
n
g


I.
UDPipe
Pre-process
II.
Feature
Extraction
III.
Unlabeled
Parsing
IV.
Arc
Labeling
Universal
dependency
parsing
Bi-directional LSTM
Universal
dependency
parsing

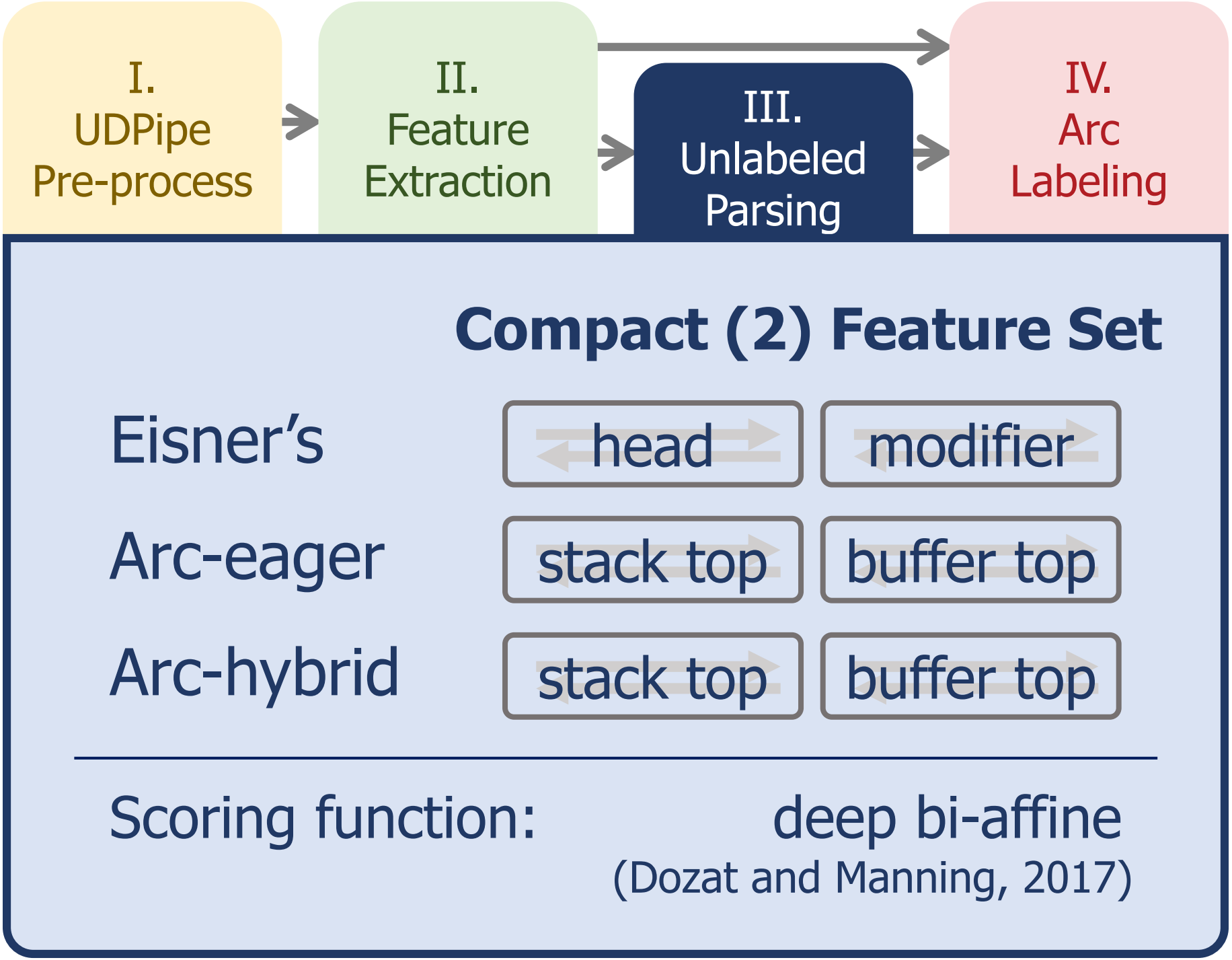

I.
UDPipe
Pre-process
II.
Feature
Extraction
III.
Unlabeled
Parsing
IV.
Arc
Labeling
Reparsing by Eisner's
(Sagae and Lavie, 2006)
Eisner's
Arc-eager
Global
NEW
!!!
Arc-hybrid
Global
Bi-LSTM features



Global Transition-based Parsing

- $O(n^3)$ *Exact* decoders
- Arc-eager and Arc-hybrid systems
- Large-margin *global* training
- Dynamic programming (Huang and Sagae, 2010; Kuhlmann, Gómez-Rodríguez and Satta, 2011)

* Shi, Huang and Lee (2017, EMNLP)


I.
UDPipe
Pre-process
II.
Feature
Extraction
III.
Unlabeled
Parsing
IV.
Arc
Labeling
Compact (2) Feature Set
Eisner's
head
modifier
Arc-eager
stack top
buffer top
Arc-hybrid
stack top
buffer top
Scoring function:
deep bi-affine
(Dozat and Manning, 2017)

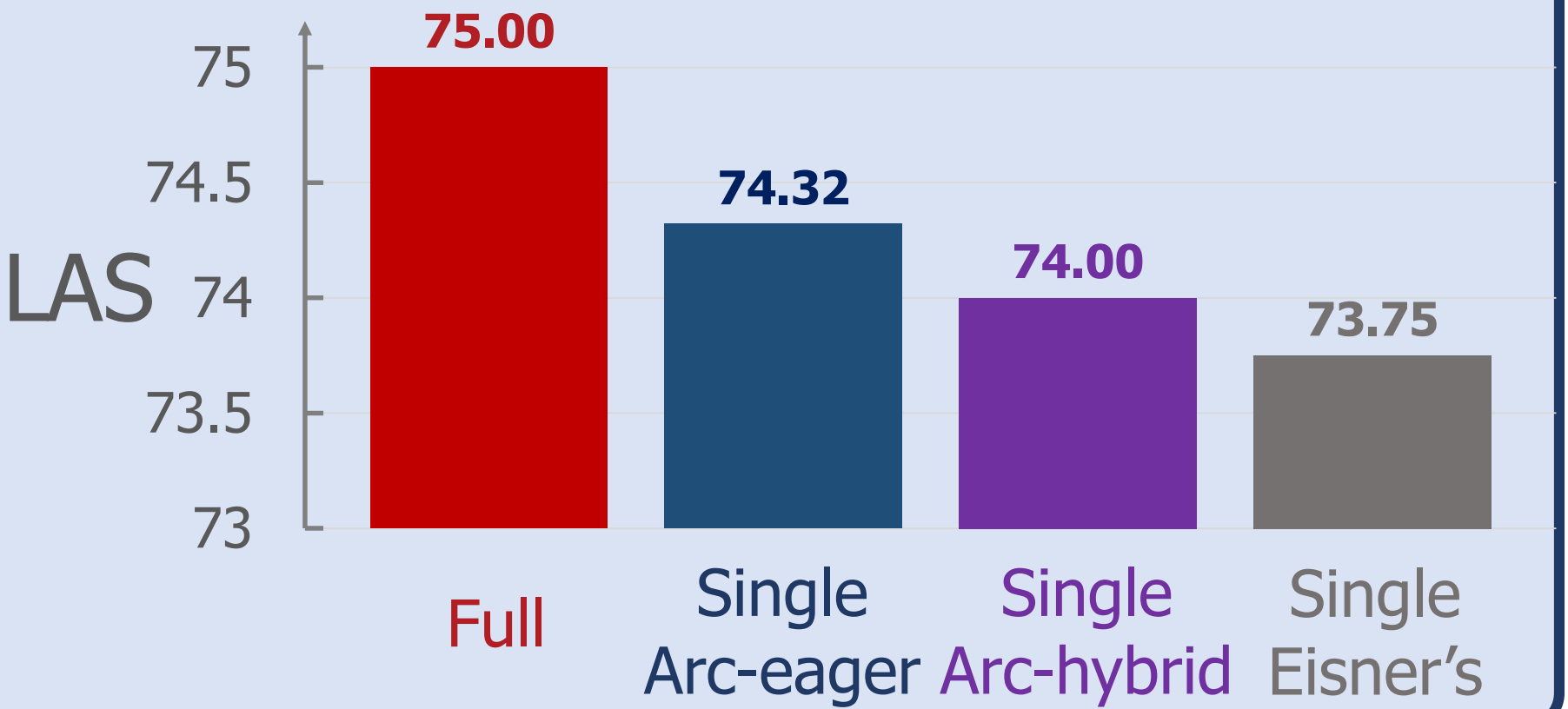
I. UDPipe Pre-process

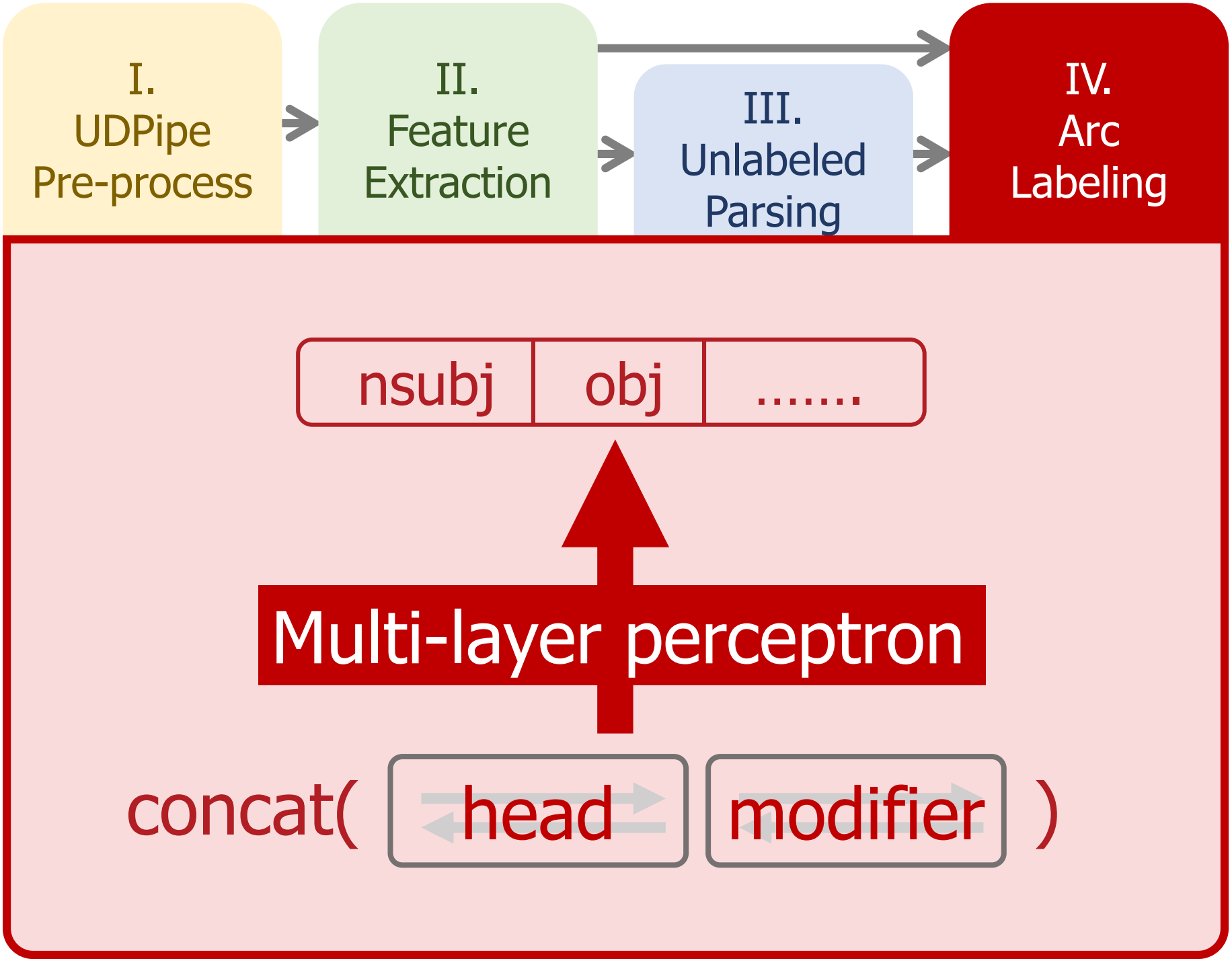
II. Feature Extraction

III. Unlabeled Parsing

IV. Arc Labeling

Ensembling




I.
UDPipe
Pre-process
II.
Feature
Extraction
III.
Unlabeled
Parsing
IV.
Arc
Labeling
nsubj
obj
.....
Multi-layer perceptron
concat(
head
modifier
)

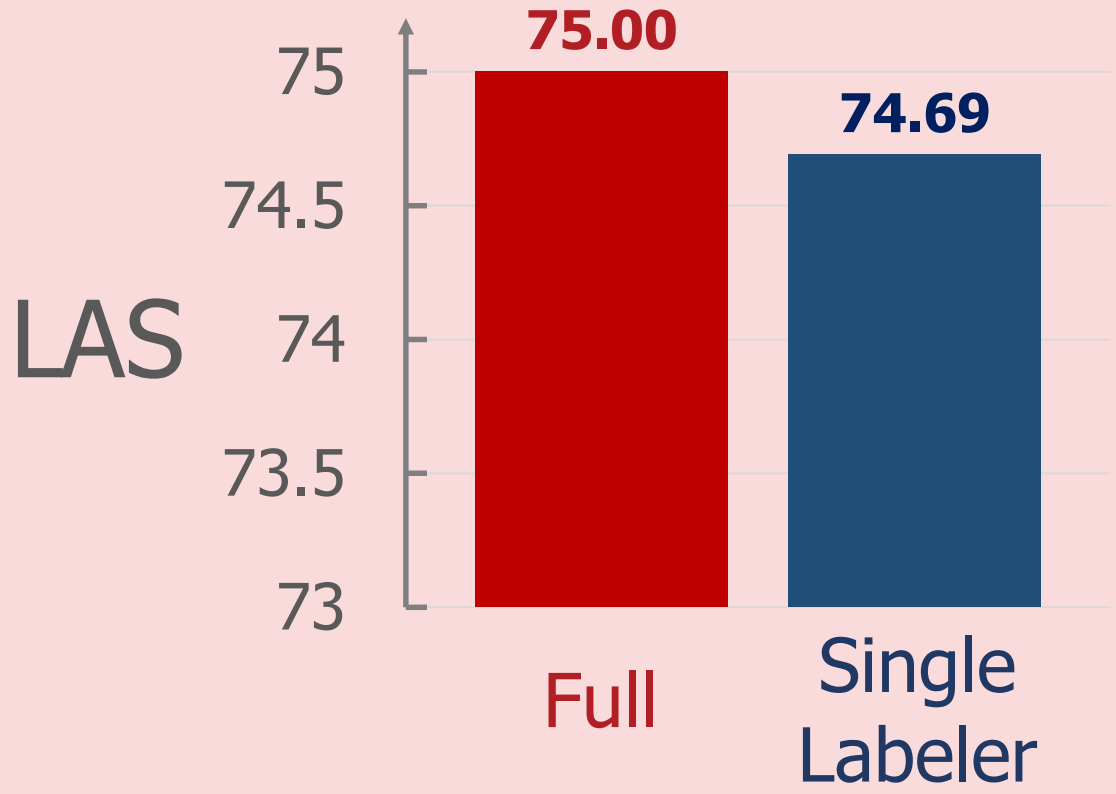
I. UDPipe Pre-process

II. Feature Extraction

III. Unlabeled Parsing

IV. Arc Labeling

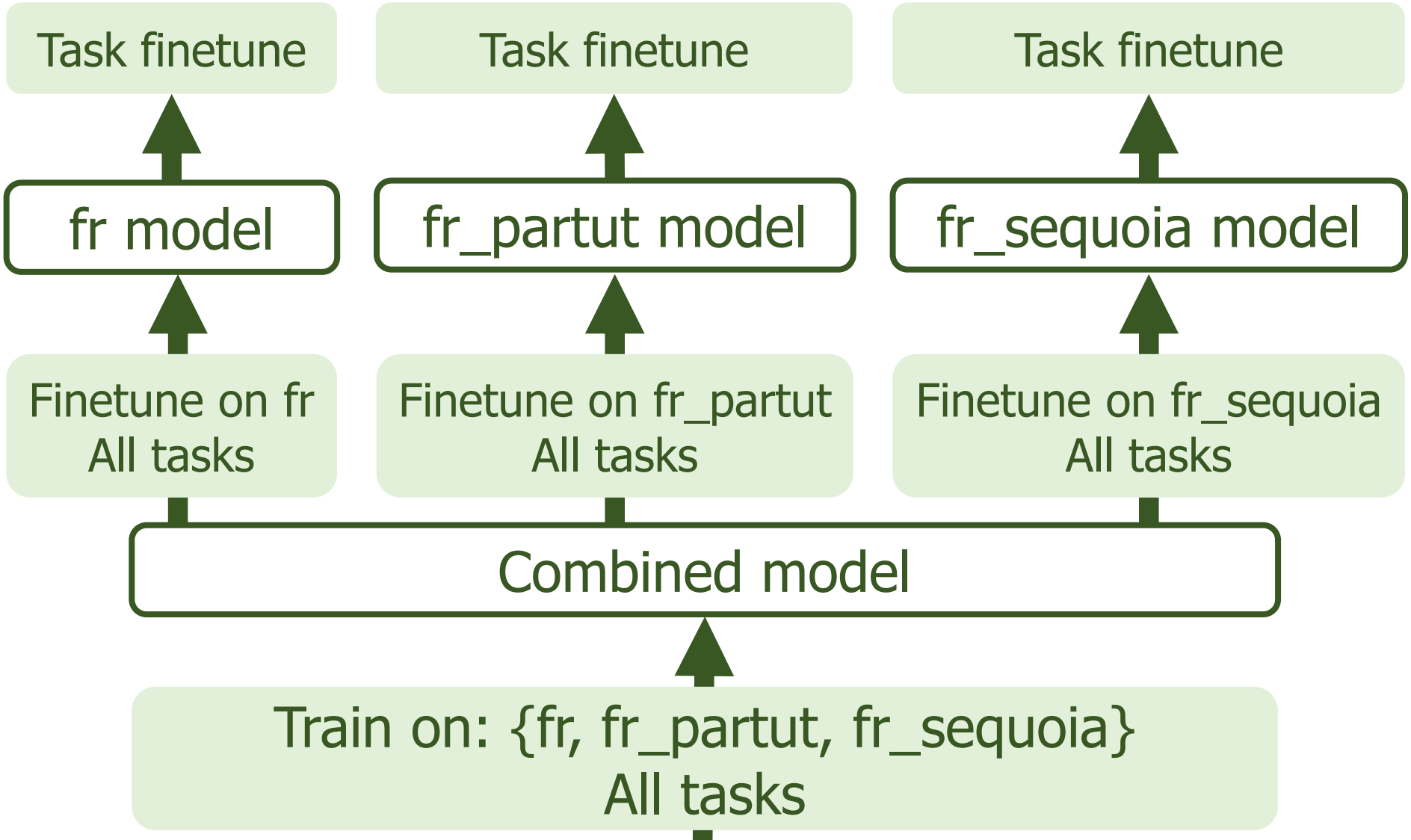
Effect of Ensemble



Results — Official Ranking

Big Treebanks	2
Small Treebanks	1
PUD Treebanks	2
Surprise Languages	1
Overall	2

Strategies — Small Treebanks



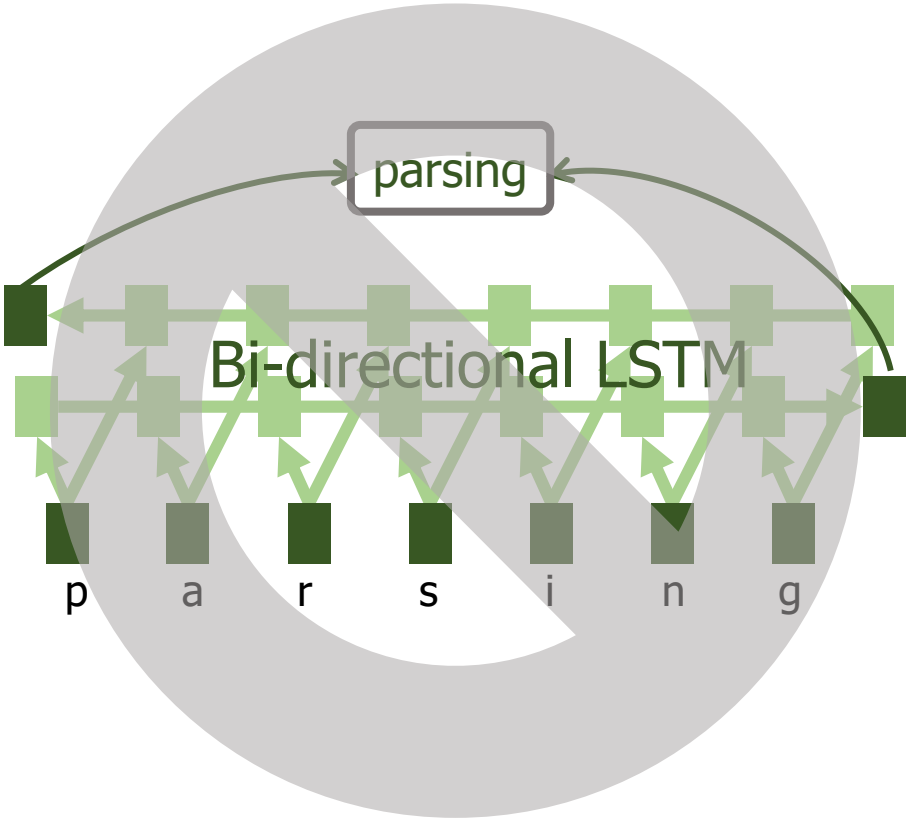
Results — Small Treebanks

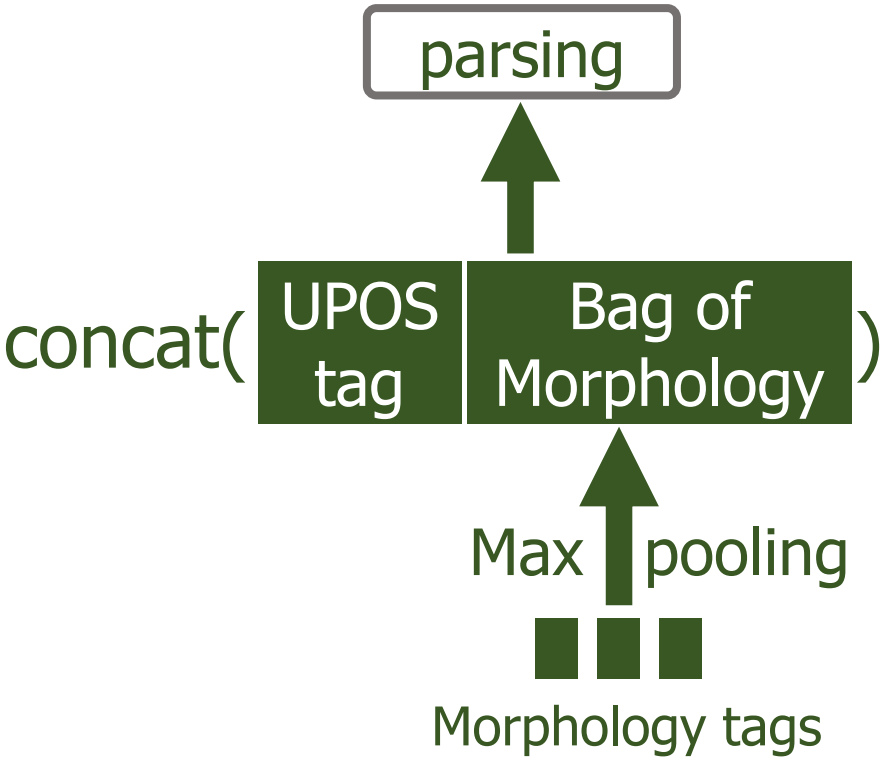
Train Treebank	Test Treebank: fr	Test Treebank: fr_partut	Test Treebank: fr_sequoia
fr	84.09		
fr_partut		79.53	
fr_sequoia			84.65
Combined	87.57	85.57	82.80
+Finetune	**87.87**	**86.65**	**86.37**

* UAS results on dev set, using gold segmentation

Strategies — Surprise Languages

- Train on a source language (selected via WALS)
- Delexicalized parser





Results — Surprise Languages

Target	Source*	Ranking
Buryat	Hindi	2
Upper Sorbian	Czech	1
Kurmanji	Persian	1
North Sámi	Finnish	1
Average		1

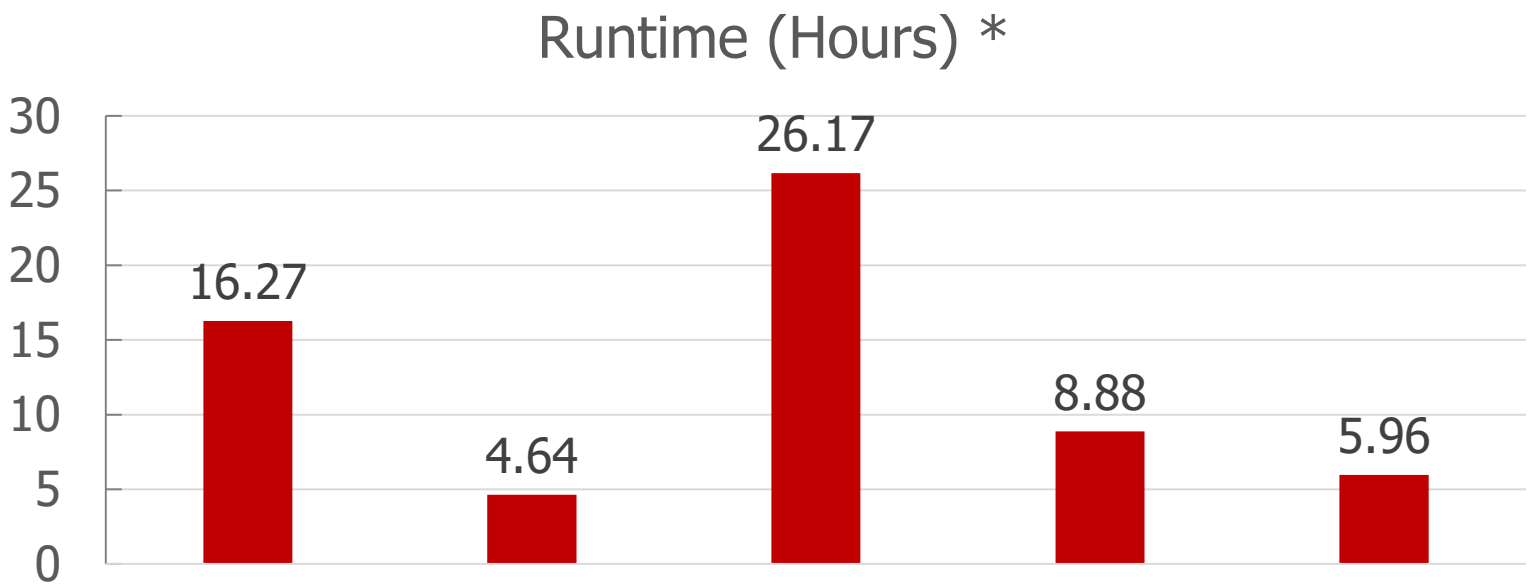
*selected via WALS

Implementation

- Neural networks — dy/net
- Parsing algorithms — Cython
- Hardware — 

X 2
- Training time — Approx. 1 week

Efficiency



	Stanford (Stanford)	C2L2 (Ithaca)	IMS (Stuttgart)	HIT-SCIR (Harbin)	LATTICE (Paris)
LAS	76.30	75.00	74.42	72.11	70.93
CPUs	4	2	12	1	8
RAM	16	8	64	8	32

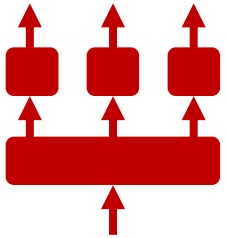
* *Not* Benchmark Results

Combining Global Models for Parsing Universal Dependencies

$$\operatorname*{argmax}_{y \in \mathcal{Y}}$$

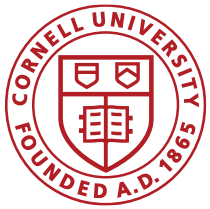
- Global transition-based models

- Ensemble



- Two-stage fine-tuning

https://github.com/CoNLL-UD-2017/C2L2



Team C2L2 —

Tianze Shi, Felix G. Wu, Xilun Chen, Yao Cheng