Nested Named Entity Recognition as Latent Lexicalized Constituency Parsing

Chao Lou, Songlin Yang, Kewei Tu

School of Information Science and Technology, ShanghaiTech University

Task & Contribution

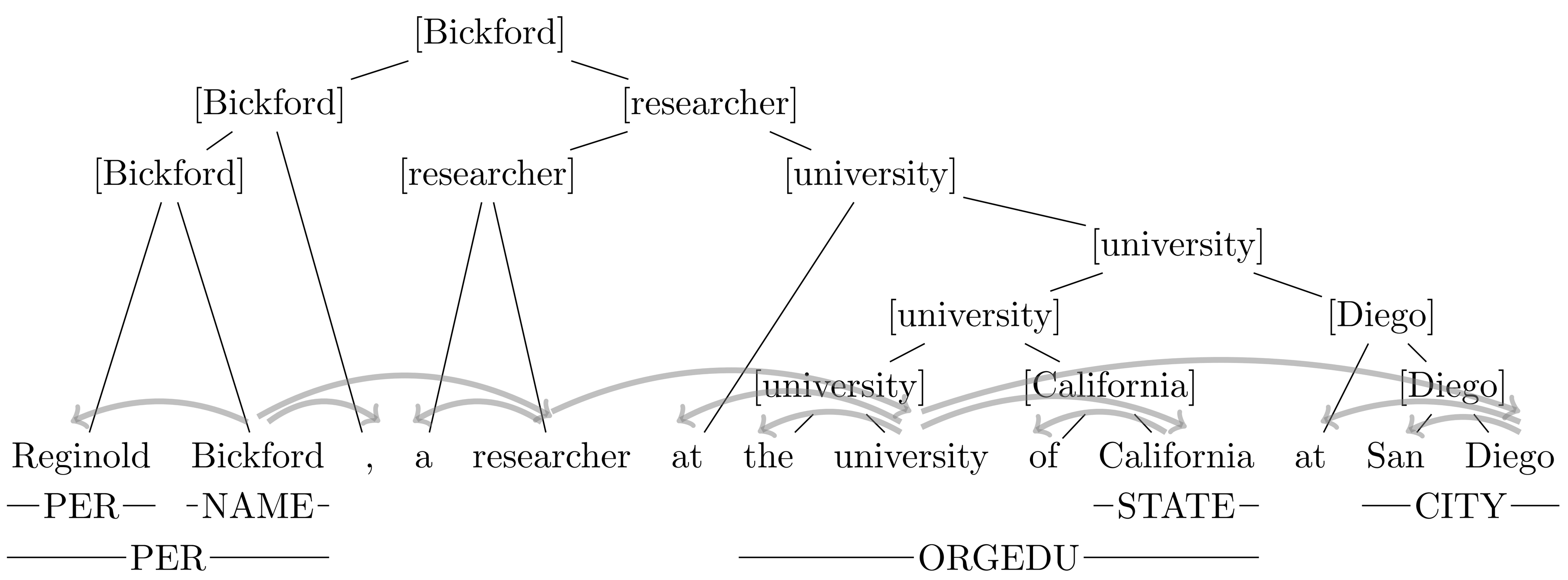
Nested NER brings more flexibility than flat NER by allowing nested structures.

An observation: **entity heads** are important clues for entity recognition.

We formulate nested NER as latent lexicalized constituency parsing, a probabilistically principled method that enables **exact global inference**, meanwhile taking **entity heads** into accounts.

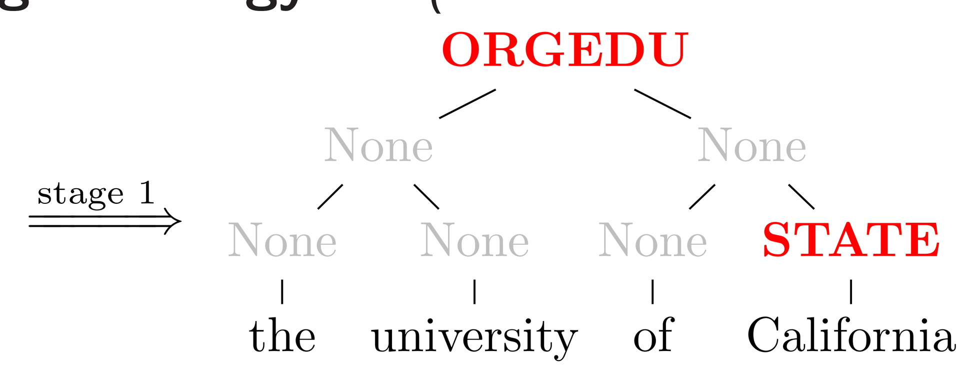
- The training algorithm handles unobserved constituents and entity heads.
- **A head regularization loss** that biases the prediction of entity heads to be distinct from each other.
- **A head-aware labeling mechanism** that types entities considering uncertain entity heads.

An example lexicalized constituency tree. Inner nodes are labeled with span heads. Arcs represent the dependency tree read from span heads. Annotations under the sentence are nested entities, or **a partially-observed tree**.

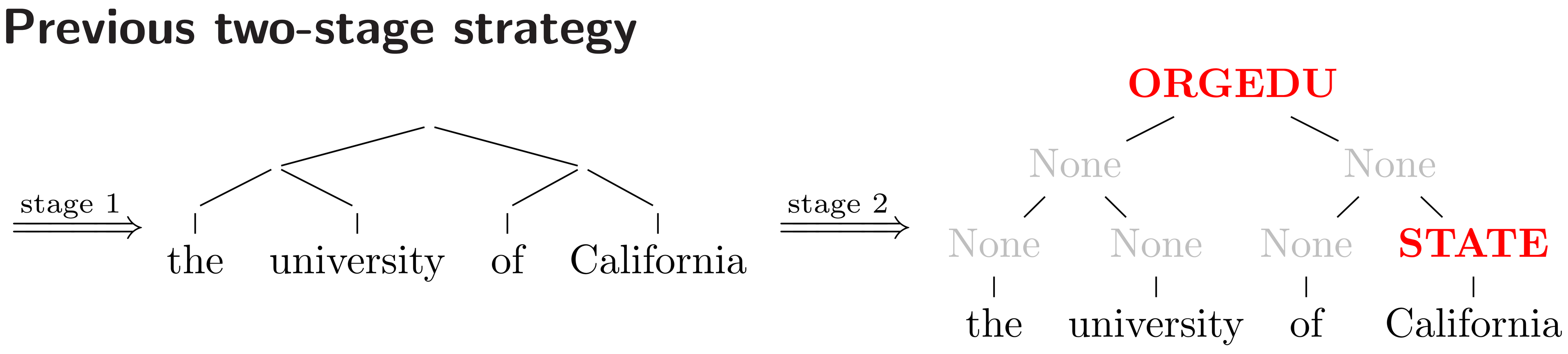


Method

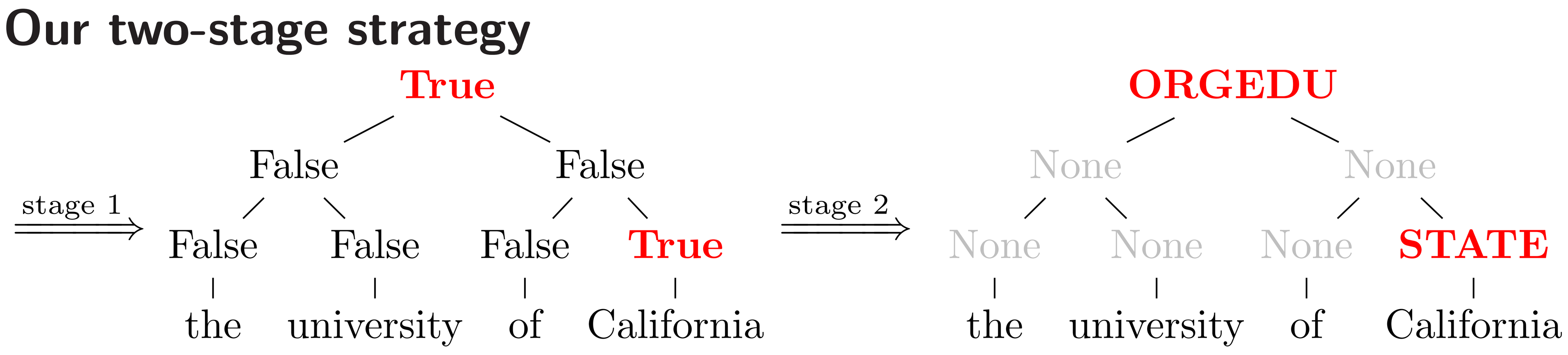
- **Previous one-stage strategy** *(Heads are omitted for brevity)*



- **Previous two-stage strategy**



- **Our two-stage strategy**



  - In stage 1, the model is asked to distinguish observed entities (**True**) and unobserved constituents (False). More efficient training.
  - In stage 2, the model no longer suffers from the label imbalance problem caused by the empty label (None).
- **Head regularization**
  Hardly or softly constain the model not to assign the same head to entities. The regularization term is the KL divergence of parse tree distributions with/without the constraint.
- **Head-aware labeling**
  The labeler takes in both span representations and head representations to predict entity labels. For span (i, j) and the possible head k, we take the expected loss under the marginal distribution $p(k|i, j)$ as the labeling loss.

Results

Model	ACE2004 P	ACE2004 R	ACE2004 F1
Pyramid-Basic	86.08	86.48	86.28
W(max)	86.27	85.09	85.68
PO-TreeCRFs	87.62	87.57	87.60
Seq2set	87.05	86.26	86.65
Locate&Label	87.27	86.61	86.94
BARTNER	87.27	86.41	86.84
Ours	87.39	88.40	87.90

Model	ACE2005 P	ACE2005 R	ACE2005 F1
SH	83.30	84.69	83.99
Pyramid-Basic	83.95	85.39	84.66
W(max)	85.28	84.15	84.71
PO-TreeCRFs	83.34	85.67	84.49
Seq2set	83.92	84.75	84.33
Locate&Label	86.02	85.62	85.82
BARTNER	83.16	86.38	84.74
Ours	85.97	87.87	86.91

Model	GENIA P	GENIA R	GENIA F1
SH	77.46	76.65	77.05
Pyramid-Basic	78.45	78.94	79.19
W(max)	79.20	78.16	78.67
PO-TreeCRFs	79.10	76.53	77.80
Seq2set	78.33	76.66	77.48
Locate&Label	76.80	79.02	77.89
BARTNER	78.57	79.3	78.93
Ours	78.39	78.50	78.44

Model	NNE P	NNE R	NNE F1
Pyramid-Basic	93.97	94.79	94.37
Ours	94.32	94.97	94.64

Table 1. Results on ACE2004, ACE2005, GENIA and NNE test sets.

Model	P	R	F1
Unstructured(1-stage)	83.76	87.17	85.43
Unstructured(2-stage)	84.23	86.62	85.41
1-stage	84.08	87.52	85.76
1-stage + LEX	84.26	87.83	86.01
2-stage	84.68	87.33	85.99
2-stage + LEX	84.60	87.80	86.17
2-stage (0-1) + LEX	84.83	87.87	86.32
- parsing	84.26	87.40	85.83
+ head regularization	85.84	87.30	86.56
+ head-aware labeling	85.50	87.77	86.62
+ both (our final model)	85.97	87.87	86.91

Table 2. Ablation study.