Enhanced Representations and Efficient Analysis of Syntactic Dependencies Within and Beyond Tree Structures

Tianze Shi
Syntactic Analysis

• Frances McDormand plays Fern in “Nomadland”.

• Joe Biden won the 2020 presidential election.

• The 2020 Summer Olympics will begin on Friday.
• The 2020 Summer Olympics will begin on Friday.
• Joe Biden won the 2020 presidential election.
• Frances McDormand plays Fern in "Nomadland".
Syntactic Analysis

When in the Course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the powers of the earth, the separate and equal station to which the Laws of Nature and of Nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation.

(The opening sentence of the *Declaration of Independence*)
When in the Course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the powers of the earth, the separate and equal station to which the Laws of Nature and of Nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation.

(The opening sentence of the *Declaration of Independence*)
Dependency Trees

- Each word is a node
- Directed edges represent asymmetric relations
- Spanning tree over the nodes

I like syntactic parsing
Dependency Parsing

\[
\text{argmax}_{y \in \mathcal{Y}} \text{score}_\theta(y|x)
\]
Dependency Parsing

\[
\arg\max_{y \in \mathcal{Y}} \text{score}_\theta(y|x)
\]
Dependency Parsing

\[
\text{argmax}_{y \in \mathcal{Y}} \text{score}_\theta (y|x)
\]
Dependency Parsing

• Output space
  • All possible spanning trees over the sentence
Number and Types of Core Arguments

• An example from the winning system at the CoNLL 2017 shared task

System Prediction

Gold Standard

How come no one bothers to ask ...
Dependency Parsing

• Output space
  • All possible spanning trees over the sentence

• Common evaluation metrics
  • Unlabeled attachment score (UAS)
  • Labeled attachment score (LAS)

• Models are typically trained to minimize
  • (Individual) Attachment errors
  • (Individual) Labeling errors
# Universal Dependencies Taxonomy

<table>
<thead>
<tr>
<th>Core arguments</th>
<th>Nominals</th>
<th>Clauses</th>
<th>Modifier words</th>
<th>Function words</th>
</tr>
</thead>
<tbody>
<tr>
<td>nsubj, obj, iobj</td>
<td>csubj, ccomp, xcomp</td>
<td>advmod, discourse</td>
<td>aux, cop, mark</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-core dependents</th>
<th>Nominal dependents</th>
<th>Coordination</th>
<th>MWE</th>
<th>Loose</th>
<th>Special</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>obl, vocative, expl, dislocated</td>
<td>nmod, appos, nummod</td>
<td>conj, cc</td>
<td>fixed, flat, compound</td>
<td>list, parataxis</td>
<td>orphan, goeswith, reparandum</td>
<td>punct, root, dep</td>
</tr>
</tbody>
</table>
## Universal Dependencies Taxonomy

<table>
<thead>
<tr>
<th>Core arguments</th>
<th>Nominals</th>
<th>Clauses</th>
<th>Modifier words</th>
<th>Function words</th>
</tr>
</thead>
<tbody>
<tr>
<td>nsubj, obj, iobj</td>
<td>csubj, ccomp, xcomp</td>
<td>advcl</td>
<td>advmod, discourse</td>
<td>aux, cop, mark</td>
</tr>
<tr>
<td>Non-core dependents</td>
<td>obl, vocative, expl, dislocated</td>
<td>acl</td>
<td>amod</td>
<td>det, clf, case</td>
</tr>
<tr>
<td>Nominal dependents</td>
<td>nmod, appos, nummod</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Coordination
- conj, cc

### MWE
- fixed, flat, compound

### Loose
- list, parataxis

### Special
- orphan, goeswith, reparandum

### Others
- punct, root, dep
# Universal Dependencies Taxonomy

<table>
<thead>
<tr>
<th>Core arguments</th>
<th>Nominals</th>
<th>Clauses</th>
<th>Modifier words</th>
<th>Function words</th>
</tr>
</thead>
<tbody>
<tr>
<td>nsubj, obj, iobj</td>
<td>csubj, ccomp, xcomp</td>
<td>advcl</td>
<td>advmod, discourse</td>
<td>aux, cop, mark</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-core dependents</th>
<th>Nominal dependents</th>
<th>Coordination</th>
<th>MWE</th>
<th>Loose</th>
<th>Special</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>obl, vocative, expl, dislocated</td>
<td>nmod, appos, nummod</td>
<td>conj, cc</td>
<td>fixed, flat, compound</td>
<td>list, parataxis</td>
<td>orphan, goeswith, reparandum</td>
<td>punct, root, dep</td>
</tr>
</tbody>
</table>
This Dissertation

Scoring  Learning  Decoding
This Dissertation

Scoring  Learning  Decoding

Representation

Structural constraints

Linguistic notions
I like syntactic parsing

Augmenting Trees

Shi and Lee (EMNLP, 2018)

Shi and Lee (ACL, 2020)
Outline

Augmenting Trees

Shi and Lee (ACL, 2020)

Beyond Trees

Shi and Lee (IWPT, 2021)

Shi and Lee (ACL, 2021)
Outline

Augmenting Trees

Shi and Lee (ACL, 2020)

Beyond Trees

Shi and Lee (IWPT, 2021)

Shi and Lee (ACL, 2021)

Shi and Lee (EMNLP, 2018)
Headless Multi-Word Expressions (MWEs)

- They are frequent
  - Including named entities
    - My bank is *Wells Fargo*.
    - ACL’21 starts on *August 1, 2021*.
  - And beyond named entities
    - The candidates matched each other *insult for insult*.
      (Jackendoff, 2008)

- They show up in different representations
  - NER
  - SRL
  - Parsing
  - ...
Begin/Inside/Outside Tagging

- BIO tagging is a common solution for span extraction, e.g., NER

A monument to Martin Luther King

O O O O B I I
Headless MWEs in Treebanks

- Special relations to denote headless MWE spans
- All tokens attached to the first token – “in principle arbitrary”

(Universal Dependencies annotation guideline)

(Universal Dependencies)

(The MWE-Aware English Dependency Corpus)
Main Idea

Parsing View

Consistency

Tagging View

A monument to Martin Luther King
This Dissertation

- Scoring
- Learning
- Decoding

Structural constraints

Linguistic notions

Representation
Scoring

• Dozat and Manning (2017)’s state-of-the-art dependency parser
• + Tagging

\[ P(y|x) = \frac{1}{Z_x} \prod_{i=1}^{n} P(h_i|x_i)P(r_i|x_i, h_i)P(t_i|x_i) \]

Attachment  Relation labeling  MWE BIO Tagging
Model: Attachment Scoring

\[ P(y|x) = \frac{1}{Z_x} \prod_{i=1}^{n} P(h_i|x_i)P(r_i|x_i, h_i)P(t_i|x_i) \]

- **Input Text**
  - I
  - like
  - syntactic
  - parsing

- **Embeddings**

- **Contextualized Representations**

- **MLP\text{ATT–MOD}**

- **MOD**

- **MLP\text{ATT–HEAD}**

- **HEAD**

- **Biaffine**

- **Feature Extractor (bi-LSTM / BERT)**

- **Scoring Attachments**

- Score of attaching \( I \) to \( \text{like} \)
Model: Label Scoring

\[ P(y|x) = \frac{1}{Z_x} \prod_{i=1}^{n} P(h_i|x_i)P(r_i|x_i, h_i)P(t_i|x_i) \]

**Input Text**
- I
- like
- syntactic
- parsing

**Contextualized Representations**

**MLP^{LAB\text{-MOD}}**

**MLP^{LAB\text{-HEAD}}**

**Feature Extractor (bi-LSTM / BERT)**

**Scoring Labels**

**MOD**

**Biaffine**

**HEAD**

nsubj
Model: Tagging

\[ P(y|x) = \frac{1}{Z_x} \prod_{i=1}^{n} P(h_i|x_i)P(r_i|x_i, h_i)P(t_i|x_i) \]
Learning and Inferencing

Baseline
- Parsing Module
- Feature Extractor
- Sentence

Multi-task Learning (MTL)
- Jointly Trained
- Parsing Module
- Tagging Module
- Shared Feature Extractor
- Sentence

Joint Decoding (Enforce Consistency)
- Joint Decoder
- Parsing Module
- Tagging Module
- Shared Feature Extractor
- Sentence
Joint Decoding

- Key idea: add a deduction rule (axiom) into Eisner’s (1996) algorithm

Axioms:

- **R-INIT:** \[ \frac{\mathcal{L}}{i} : \log P(t_i = O) \]
- **L-INIT:** \[ \frac{\mathcal{L}}{i} : 0 \]

- **R-MWE:** \[ \frac{\mathcal{L}}{i \ j} : \delta(i, j), \]

where \( \delta(i, j) = \log P(t_i = B) + \sum_{k=i+1}^{j} (\log P(t_k = 1) + \log P(h_k = i)) \)

Deduction Rules:

- **R-COMB:** \[ \frac{\mathcal{L}}{i \ k} : s_1 \quad \frac{\mathcal{L}}{k \ j} : s_2 \quad \frac{\mathcal{L}}{i \ j} : s_1 + s_2 \]
- **R-LINK:** \[ \frac{\mathcal{L}}{i \ k} : s_1 \quad \frac{\mathcal{L}}{k + 1 \ j} : s_2 \quad \frac{\mathcal{L}}{i \ j} : s_1 + s_2 + \log P(h_j = i) \]
- **L-COMB:** \[ \frac{\mathcal{L}}{j \ k} : s_1 \quad \frac{\mathcal{L}}{k \ i} : s_2 \quad \frac{\mathcal{L}}{j \ i} : s_1 + s_2 \]
- **L-LINK:** \[ \frac{\mathcal{L}}{j \ k - 1} : s_1 \quad \frac{\mathcal{L}}{k \ i} : s_2 \quad \frac{\mathcal{L}}{j \ i} : s_1 + s_2 + \log P(h_j = i) \]
Experiment Results – “Standard” Parsing Metrics

- Parsing Only: 82.60
- MTL: 82.69
- Joint Decoding: 82.55
<table>
<thead>
<tr>
<th>Approach</th>
<th>F1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parsing Only</td>
<td>79.13</td>
</tr>
<tr>
<td>MTL</td>
<td>81.39</td>
</tr>
<tr>
<td>Joint Decoding</td>
<td>82.19</td>
</tr>
</tbody>
</table>
# Universal Dependencies Taxonomy

<table>
<thead>
<tr>
<th>Core arguments</th>
<th>Nominals</th>
<th>Clauses</th>
<th>Modifier words</th>
<th>Function words</th>
</tr>
</thead>
<tbody>
<tr>
<td>nsubj, obj, iobj</td>
<td>csubj, ccomp, xcomp</td>
<td>advcl</td>
<td>advmod, discourse</td>
<td>aux, cop, mark</td>
</tr>
<tr>
<td>Non-core dependents</td>
<td>obl, vocative, expl, dislocated</td>
<td>acl</td>
<td>amod</td>
<td>det, clf, case</td>
</tr>
<tr>
<td>Nominal dependents</td>
<td>nmod, appos, nummod</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coordination</th>
<th>MWE</th>
<th>Loose</th>
<th>Special</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>conj, cc</td>
<td>fixed, flat, compound</td>
<td>list, parataxis</td>
<td>orphan, goeswith, reparandum</td>
<td>punct, root, dep</td>
</tr>
</tbody>
</table>
Valency

• Valency: Type and number of dependents a word takes

(Tesnière, 1959, *inter alia*)
An Empirical Definition of Valency Patterns

- Fix a set of syntactic relations $R$, e.g., core arguments
- Encode a token’s linearly-ordered dependent relations within $R$

He says that you like to swim.
Main Idea to Incorporate Valency Patterns

He says that you like to swim.
Scoring

• Dozat and Manning (2017)’s state-of-the-art dependency parser
• + Tagging/Supertagging

\[
P(y|x) = \frac{1}{Z_x} \prod_{i=1}^{n} P(h_i|x_i) P(r_i|x_i, h_i) P(t_i|x_i)
\]

Attachment  Relation labeling  MWE/Valency Tagging
Decoding with Head Automata \cite{Eisner1999}

R-INIT: $\alpha^L \odot \alpha_1^R \odot \alpha_2^R \odot \alpha^R \quad \triangleleft \quad h \quad h$

R-COMB: $\alpha^L \odot \alpha_1^R \odot \alpha_2^R \odot \alpha^R \odot \alpha_2^R \odot \alpha^R \quad \triangleleft \quad h \quad j \quad j \quad i$

R-LINK: $\alpha^L \odot \alpha_1^R \odot a \odot \alpha_2^R \odot \alpha^R \quad \triangleleft \quad h \quad i \quad i+1 \quad j \quad h \rightarrow j$

$\alpha^L \odot \alpha_1^R \odot a \odot \alpha_2^R \odot \alpha^R \quad \triangleleft \quad h \quad i \quad i+1 \quad j \quad h \rightarrow j, r \notin \mathcal{R}$
### Experiment Results – Valency Augmented Parsing

<table>
<thead>
<tr>
<th>LAS</th>
<th>Baseline</th>
<th>Ours (Core MTL)</th>
<th>Ours (Joint Decoding)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80</td>
<td>81</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>83.6</td>
<td>83.8</td>
<td>83.9</td>
</tr>
</tbody>
</table>

**LAS**

- **Baseline**
- **Ours (Core MTL)**
- **Ours (Joint Decoding)**

**MTL** = Multi-task learning
Experiment Results – Valency Augmented Parsing

VPA = Valency pattern accuracy

MTL = Multi-task learning

VPA

<table>
<thead>
<tr>
<th>Baseline</th>
<th>Ours (Core MTL)</th>
<th>Ours (Joint Decoding)</th>
</tr>
</thead>
<tbody>
<tr>
<td>95.8</td>
<td>96.0</td>
<td>96.7</td>
</tr>
</tbody>
</table>

LAS

<table>
<thead>
<tr>
<th>Baseline</th>
<th>Ours (Core MTL)</th>
<th>Ours (Joint Decoding)</th>
</tr>
</thead>
<tbody>
<tr>
<td>83.6</td>
<td>83.8</td>
<td>83.9</td>
</tr>
</tbody>
</table>
Experiment Results – Valency Augmented Parsing

VPA = Valency pattern accuracy

MTL = Multi-task learning

<table>
<thead>
<tr>
<th>VPA</th>
<th>Baseline</th>
<th>Ours (Core MTL)</th>
<th>Ours (Joint Decoding)</th>
</tr>
</thead>
<tbody>
<tr>
<td>83.6</td>
<td>80.9</td>
<td>85.4</td>
<td></td>
</tr>
<tr>
<td>83.8</td>
<td>82.0</td>
<td>81.8</td>
<td></td>
</tr>
<tr>
<td>83.9</td>
<td>82.0</td>
<td>82.0</td>
<td></td>
</tr>
<tr>
<td>85.4</td>
<td>82.0</td>
<td>82.0</td>
<td></td>
</tr>
</tbody>
</table>

95.8 96.0 96.7

95.5 96.0 96.5 97.0
Outline

Augmenting Trees

Shi and Lee (ACL, 2020)

Beyond Trees

Shi and Lee (IWPT, 2021)

Shi and Lee (ACL, 2021)
Outline

Augmenting Trees

Beyond Trees

Shi and Lee (ACL, 2020)

Shi and Lee (IWPT, 2021)

Shi and Lee (ACL, 2021)
Coordination in Dependency Structures

Constraints on Coordination

Constraints on Coordination

Icons credited to Gregor Cresnar, Jacob Halton, Muhammad Auns, and Umer Younas (CC-BY licensed)
Coordination in Dependency Structures

hot coffee or tea

Icons credited to Gregor Cresnar, Jacob Halton, Muhammad Auns, and Umer Younas (CC-BY licensed)
Coordination in Dependency Structures

I prefer hot coffee or tea and a croissant.

Icons credited to Gregor Cresnar, Jacob Halton, Muhammad Auns, and Umer Younas (CC-BY licensed)
Coordination in Dependency Structures

I prefer **hot coffee** or **tea** and a **croissant**

Icons credited to Gregor Cresnar, Jacob Halton, Muhammad Auns, and Umer Younas (CC-BY licensed)
Coordination is Difficult to Represent

• Symmetry among conjuncts

Coordination Structures in Dependency Treebanks

Martin Popel, David Mareček, Jan Štěpánek, Daniel Zeman, Zdeněk Žabokrtský
Charles University in Prague, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics (ÚFAL)
Malostranské náměstí 25, CZ-11800 Praha, Czechia

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 517–527,
Sofia, Bulgaria, August 4-9 2013. ©2013 Association for Computational Linguistics

In both cases, a number of decisions have to be made during the construction or conversion of a dependency treebank. The traditional notion of dependency does not always provide unambiguous solutions, e.g. when it comes to attaching functional words. Worse, dependency representation is at a loss when it comes to representing paratactic linguistic phenomena such as coordination, whose nature is symmetric (two or more conjuncts play the same role), as opposed to the head-modifier asymmetry of dependencies.¹

¹We use the term modifier (or child) for all types of dependent nodes including arguments.
Dependency-based Solutions

• Prague-style dependencies with coordinators as subtree roots
  (Hajič et al., 2001, 2006, 2020)

I prefer hot coffee or tea and a croissant
Dependency-based Solutions

- Enhanced UD Graphs (Schuster and Manning, 2016; Nivre et al., 2018; Bouma et al., 2020)

I prefer hot coffee or tea and a croissant
Dependency-based Solutions

- Enhanced UD Graphs (Schuster and Manning, 2016; Nivre et al., 2018; Bouma et al., 2020)

IWPT 2020 and 2021 Shared Task

I prefer **hot** coffee **or** tea **and** a croissant
## IWPT 2021 Shared Task Official Evaluation

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>ELAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>TGIF</td>
<td>89.24</td>
</tr>
<tr>
<td>2.</td>
<td>SHANGAITECH</td>
<td>87.07</td>
</tr>
<tr>
<td>3.</td>
<td>ROBERTNLP</td>
<td>86.97</td>
</tr>
<tr>
<td>4.</td>
<td>COMBO</td>
<td>83.79</td>
</tr>
<tr>
<td>5.</td>
<td>UNIPI</td>
<td>83.64</td>
</tr>
<tr>
<td>6.</td>
<td>DCU EPFL</td>
<td>83.57</td>
</tr>
<tr>
<td>7.</td>
<td>GREW</td>
<td>81.58</td>
</tr>
<tr>
<td>8.</td>
<td>FASTPARSE</td>
<td>65.81</td>
</tr>
<tr>
<td>9.</td>
<td>NUIG</td>
<td>30.03</td>
</tr>
</tbody>
</table>

**Average:** 89.24

### Best ELAS on 16/17 languages

<table>
<thead>
<tr>
<th>Language</th>
<th>ELAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic</td>
<td>81.23</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>93.63</td>
</tr>
<tr>
<td>Czech</td>
<td>92.24</td>
</tr>
<tr>
<td>Dutch</td>
<td>91.78</td>
</tr>
<tr>
<td>English</td>
<td>88.19</td>
</tr>
<tr>
<td>Estonian</td>
<td>88.38</td>
</tr>
<tr>
<td>Finnish</td>
<td>91.75</td>
</tr>
<tr>
<td>French</td>
<td>91.63</td>
</tr>
<tr>
<td>Italian</td>
<td>93.31</td>
</tr>
<tr>
<td>Latvian</td>
<td>90.23</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>86.06</td>
</tr>
<tr>
<td>Polish</td>
<td>91.46</td>
</tr>
<tr>
<td>Russian</td>
<td>94.01</td>
</tr>
<tr>
<td>Slovak</td>
<td>94.96</td>
</tr>
<tr>
<td>Swedish</td>
<td>89.90</td>
</tr>
<tr>
<td>Tamil</td>
<td>65.58</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>92.78</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>89.24</strong></td>
</tr>
</tbody>
</table>
System Overview

- **Tokenizer**
- **Sentence Splitter**
- **Multi-Word Token Expander**
- **Lemma Dictionary**
- **EUD Parser**

**Training Strategy:**
Two-Stage Finetuning
System Overview

Training Strategy: Two-Stage Finetuning

[Diagram showing the flow of data processing steps: Tokenizer → Multi-Word Token Expander → Sentence Splitter → EUD Parser → Lemma Dictionary]
TGIF: **Tree-Graph Integrated-Format** Parser

- Inspired by He and Choi *(IWPT Shared Task, 2020)*

![Diagram](image-url)

- **EUD Parser**
- **Biaffine Tree Parser**
- **Biaffine Graph Parser**
- **Relation Labeler**
TGIF: **Tree-Graph Integrated-Format Parser**

- Every connected graph must have a spanning tree
EUD Parsing Results

• Overall, +0.10% ELAS with tree-graph integration method
• Improvement on 12/17 languages

Bulgarian, Czech, English, Finnish, French, Italian, Lithuanian, Polish, Russian, Slovak, Swedish, Tamil

Arabic, Dutch, Estonian, Latvian, Ukrainian

Tree-Graph integrated method wins
Graph-only method wins
EUD Graphs

- Modifier/argument sharing
- Other phenomena (e.g., relative clauses)
- Nested coordination
- Symmetry among conjuncts

I prefer hot coffee or tea and a croissant
Looking for Other Solutions ...

Unit Coordination and Gapping in Dependency Theory

Vincenzo Lombardo and Leonardo Lesmo
Dipartimento di Informatica and Centro di Scienza Cognitiva
Università’ di Torino
c.so Svizzera 185 - 10149 Torino - Italy

Processing of Dependency-Based Grammars
(Workshop, 1998)

Dependency paradigms exhibit obvious difficulties with coordination because, differently from most linguistic structures, it is not possible to characterize the coordination construct with a general schema involving a head and some modifiers of it. The conjunction itself, has distributional properties that have nothing to do with the whole coordination. Hudson (1990, following Tesniere 1959) gives up the idea of providing a dependency structure for the coordination, and characterizes conjuncts as word strings. Conjuncts are internally organized as (possibly disconnected) dependency structures and each conjunct root is dependency related to some element of the sentence which is external to the coordination.
Adding Coordination Boundaries

I prefer (hot [coffee or tea]) and a croissant.

Icons credited to Gregor Cresnar, Jacob Halton, Muhammad Auns, and Umer Younas (CC-BY licensed)
Adding Coordination Boundaries

I prefer hot coffee or tea and a croissant.

Icons credited to Gregor Cresnar, Jacob Halton, Muhammad Auns, and Umer Younas (CC-BY licensed)
I prefer **hot (coffee or tea) and a croissant**.
Transition-based Parsing
Bubble-Hybrid Transition System

• Based on Arc-Hybrid (Kuhlmann et al., 2010)

• 6 transitions

SHIFT  LEFTARC  RIGHTARC  BUBBLEOPEN  BUBBLEATTACH  BUBBLECLOSE

Same as Arc-Hybrid  NEW
Bubble-Hybrid Transition System

\[ \text{SHIFT} \]

Stack \hspace{1cm} Buffer

\[ \ldots \hspace{1cm} b_1 \ldots \]

\[ \ldots b_1 \hspace{1cm} \ldots \]
Bubble-Hybrid Transition System

LEFTARC_{lbl}

Stack

Buffer

... s_1

b_1 ...

...}

lbl

b_1 ...

s_1
Bubble-Hybrid Transition System

\[ \text{RIGHTARC}_{\text{lbl}} \]

Diagram showing elements labeled \( s_1 \) and \( s_2 \) in a stack and buffer context.
Bubble-Hybrid Transition System

\[ \text{BubbleOPEN}_\text{lbl} \]
Bubble-Hybrid Transition System
Bubble-Hybrid Transition System

\[ \text{Stack} \quad \text{Buffer} \]

\[ \text{\textsc{BubbleAttach}}_{\text{lbl}} \]
Walkthrough of an Example Sentence

**Stack**

... | hot coffee or tea and a croissant

**Buffer**

... hot coffee or tea and a croissant

**SHIFT * 3**

... hot coffee or tea and a croissant

**BUBBLEOPEN**

... hot coffee or tea and a croissant

**SHIFT**

... hot coffee or tea and a croissant

**BUBBLEATTACH**

... hot coffee or tea and a croissant
Walkthrough of an Example Sentence

Stack

... hot coffee or tea

Buffer

and a croissant

---

**BubbleClose**

**LeftArc**

**Shift * 2**

**BubbleOpen**

---

74
Modeling

• Follows Kiperwasser and Goldberg’s (2016) parser + Greedy Decoder
Experiment Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Teranishi et al., 2017</th>
<th>Teranishi et al., 2019</th>
<th>Ours (LSTM)</th>
<th>Ours (BERT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (PTB)</td>
<td>71.08</td>
<td>75.01</td>
<td>81.63</td>
<td>85.26</td>
</tr>
<tr>
<td>NP (PTB)</td>
<td>75.47</td>
<td>77.83</td>
<td>85.26</td>
<td>83.74</td>
</tr>
<tr>
<td>GENIA</td>
<td>65.22</td>
<td>61.22</td>
<td>67.09</td>
<td>79.18</td>
</tr>
</tbody>
</table>

Teranishi et al., 2017

Teranishi et al., 2019

Ours (LSTM)

Ours (BERT)
Summary

Syntactic Phenomenon
- Headless MWEs

Involved Relation Types
- flat

Constraint / Desired Property
- Representational constraint
- Valency patterns

Proposed Method
- Joint tagging and parsing

Output Structure
- Augmented trees

Core argument structures
- \{nsubj, obj, iobj, csubj, xcomp, ccomp\} and more

Coordination
- conj, cc

Symmetry among conjuncts;
Marked coordination boundaries
- Tree-graph integration;
Bubble parsing
- Beyond trees

I like syntactic parsing hot
- coffee
- or
- tea
Limitations and Future Work

- Non-projectivity
  - Previous work: Gómez-Rodríguez, Shi, and Lee (ACL, 2018)
    Shi, Gómez-Rodríguez, and Lee (NAACL, 2018)
Limitations and Future Work

- Alternative decoding strategies
  - Previous work:
    - Shi, Huang, and Lee (EMNLP, 2017)
    - Shi, Wu, Chen, and Cheng (CoNLL Shared Task, 2017)

Graph-based

Transition-based
Limitations and Future Work

• Extrinsic evaluation on downstream tasks
  • Previous work: Shi, Malioutov, and İrsoy (EMNLP, 2020)
### Universal Dependencies Taxonomy

<table>
<thead>
<tr>
<th></th>
<th>Nominals</th>
<th>Clauses</th>
<th>Modifier words</th>
<th>Function words</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Core arguments</strong></td>
<td>nsubj, obj, iobj</td>
<td>csubj, ccomp, xcomp</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-core dependents</strong></td>
<td>obl, vocative, expl, dislocated</td>
<td>advcl</td>
<td>advmod, discourse</td>
<td>aux, cop, mark</td>
</tr>
<tr>
<td><strong>Nominal dependents</strong></td>
<td>nmod, appos, nummod</td>
<td>acl</td>
<td>amod</td>
<td>det, clf, case</td>
</tr>
</tbody>
</table>

#### Coordination
- **MWE**
  - conj, cc
  - fixed, flat, compound

#### Loose
- list, parataxis

#### Special
- orphan, goeswith, reparandum

#### Others
- punct, root, dep
All About Parsing

- Annotation
- Application
- Coverage
- Parse Trees
- Syntactic Phenomena
- Evaluation
- Multilinguality
- Computational Modeling