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.

Syntactic Analysis

- Frances McDormand *plays* Fern in "Nomadland".

$\underbrace{\text{Frances McDormand}}_{\text{subject}}$ *plays* $\underbrace{\text{Fern}}_{\text{object}}$ $\underbrace{\text{in "Nomadland"}}_{\text{modifier}}$.

plays'(Frances McDormand', Fern', in "Nomadland"')

- Joe Biden *won* the 2020 presidential election.

$\underbrace{\text{Joe Biden}}_{\text{subject}}$ *won* $\underbrace{\text{the 2020 presidential election}}_{\text{object}}$.

won'(Joe Biden', the 2020 presidential election')

- The 2020 Summer Olympics *will begin* on Friday.

$\underbrace{\text{The 2020 Summer Olympics}}_{\text{subject}}$ *will begin* $\underbrace{\text{on Friday}}_{\text{modifier}}$.

will begin'(the 2020 Summer Olympics', on Friday')

(Reddy et al., 2017)

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.

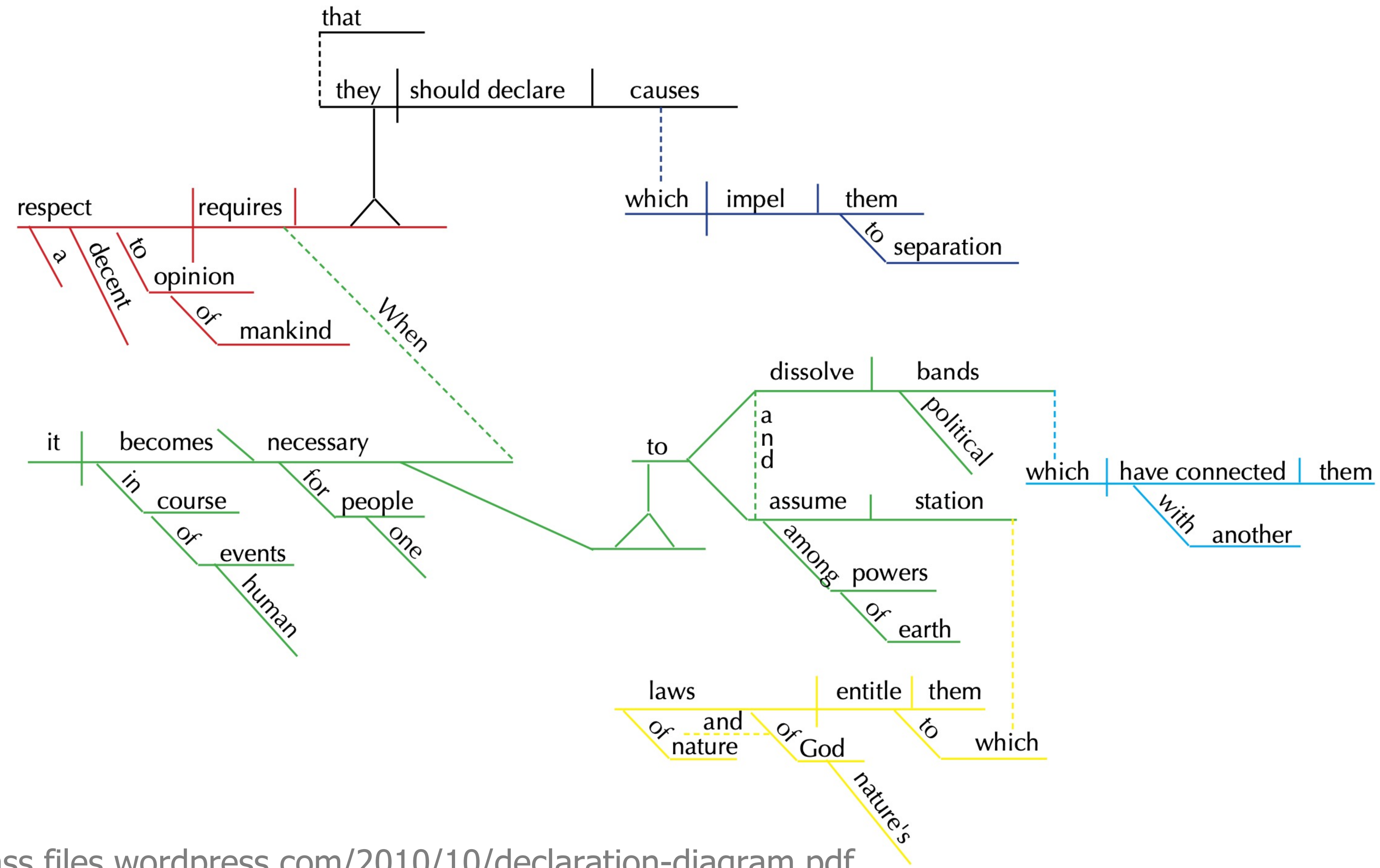
(The opening sentence of the *Declaration of Independence*)

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.

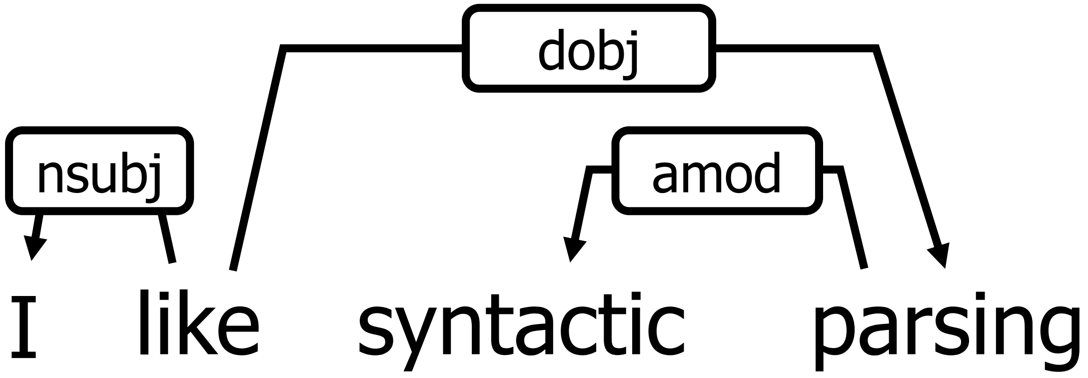
(The opening sentence of the *Declaration of Independence*)

Syntactic Analysis



Source: https://martinsclass.files.wordpress.com/2010/10/declaration-diagram.pdf

Dependency Trees



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

Dependency Parsing

candidate output dependency tree

input text

$$\underset{y \in \mathcal{Y}}{\operatorname{argmax}}\ \mathrm{score}_{\theta}(y|x)$$

Dependency Parsing

decoding

candidate output
dependency tree

input text

$$\underset{y \in \mathcal{Y}}{\operatorname{argmax}}\ \mathrm{score}_{\theta}(y|x)$$

scoring

learning

Dependency Parsing

candidate output
dependency tree

decoding

input text

$$\underset{y \in \mathcal{Y}}{\operatorname{argmax}}\ \operatorname{score}_{\theta}(y|x)$$

output
space

scoring

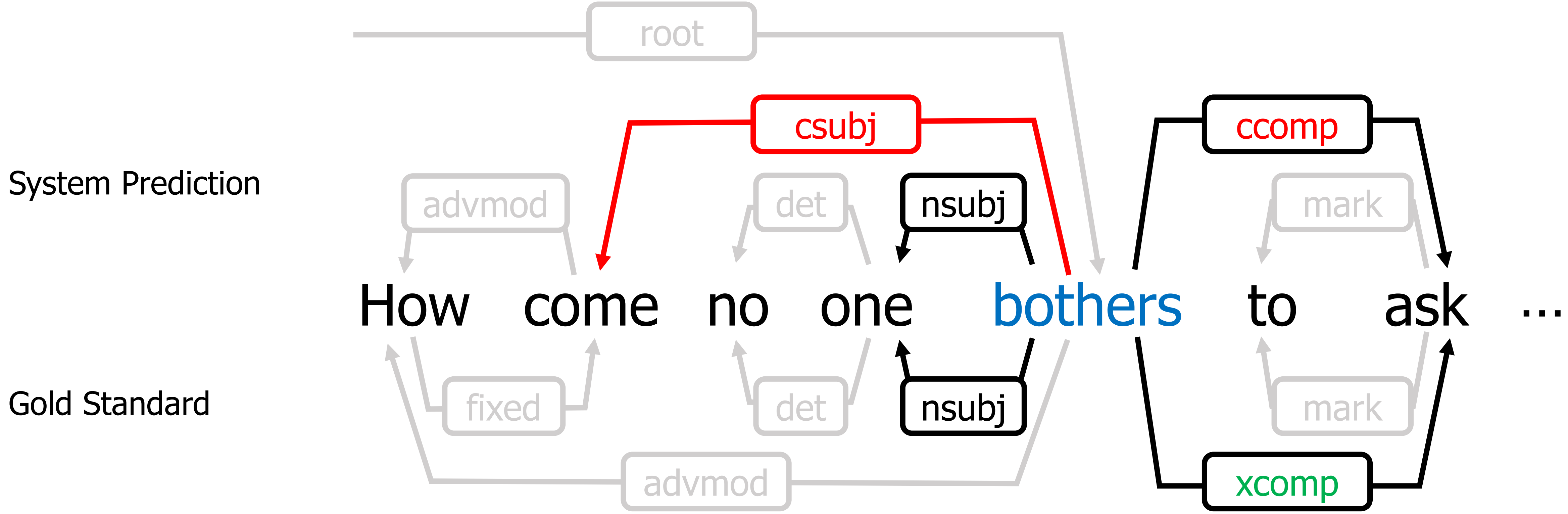
learning

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



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

	Nominals	Clauses	Modifier words	Function words
Core arguments	nsubj, obj, iobj	csubj, ccomp, xcomp		
Non-core dependents	obl, vocative, expl, dislocated	advcl	advmod, discourse	aux, cop, mark
Nominal dependents	nmod, appos, nummod	acl	amod	det, clf, case
Coordination	**MWE**	**Loose**	**Special**	**Others**
conj, cc	fixed, flat, compound	list, parataxis	orphan, goeswith, reparandum	punct, root, dep

Universal Dependencies Taxonomy

	Nominals	Clauses	Modifier words	Function words
Core arguments	nsubj, obj, iobj	csubj, ccomp, xcomp		
Non-core dependents	obl, vocative, expl, dislocated	advcl	advmod, discourse	aux, cop, mark
Nominal dependents	nmod, appos, nummod	acl	amod	det, clf, case
Coordination	**MWE**	**Loose**	**Special**	**Others**
conj, cc	fixed, flat, compound	list, parataxis	orphan, goeswith, reparandum	punct, root, dep

Universal Dependencies Taxonomy

	Nominals	Clauses	Modifier words	Function words
Core arguments	nsubj, obj, iobj	csubj, ccomp, xcomp		
Non-core dependents	obl, vocative, expl, dislocated	advcl	advmod, discourse	aux, cop, mark
Nominal dependents	nmod, appos, nummod	acl	amod	det, clf, case
Coordination	**MWE**	**Loose**	**Special**	**Others**
conj, cc	fixed, flat, compound	list, parataxis	orphan, goeswith, reparandum	punct, root, dep

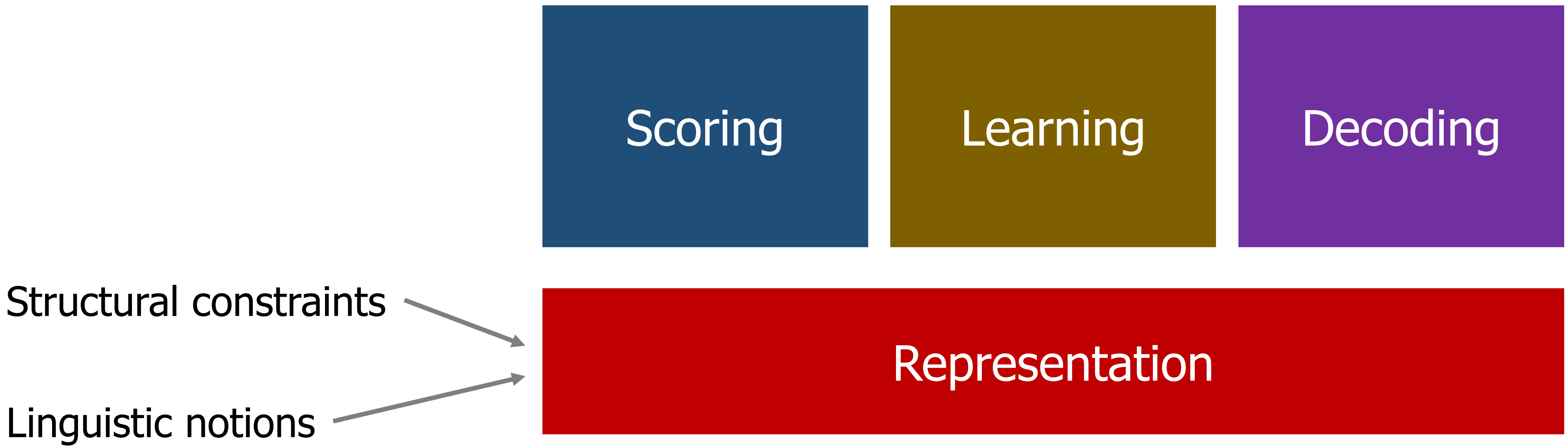
This Dissertation

Scoring

Learning

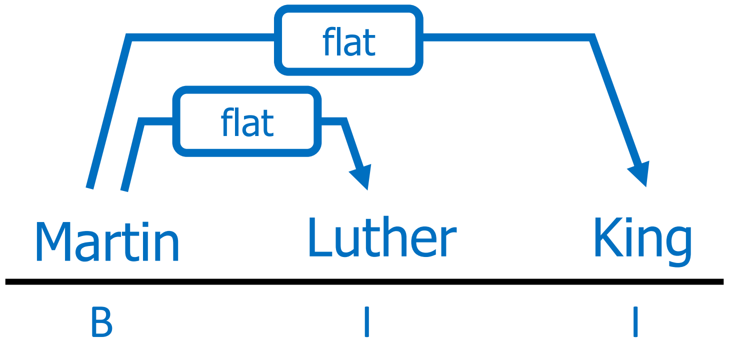
Decoding

This Dissertation

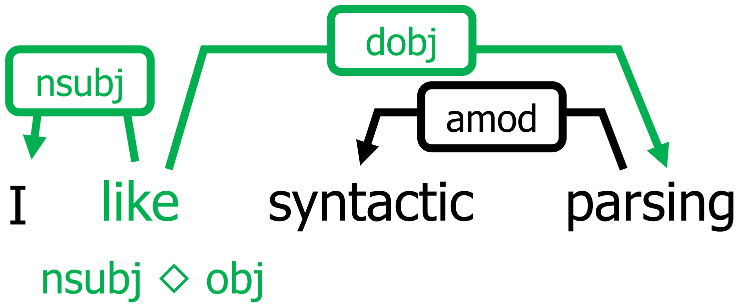


Outline

Augmenting Trees



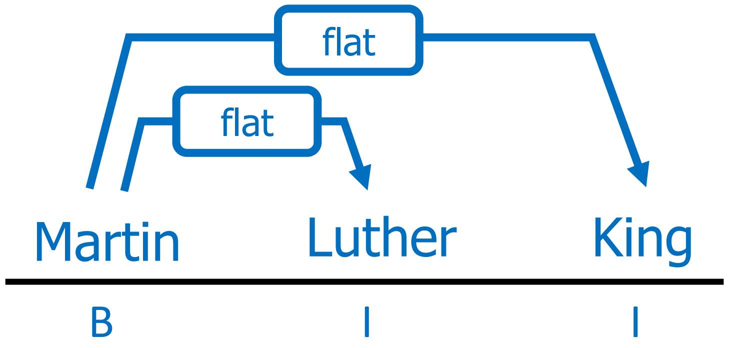
Shi and Lee (ACL, 2020)



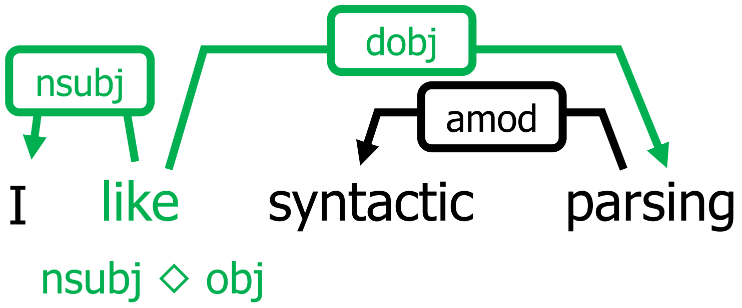
Shi and Lee (EMNLP, 2018)

Outline

Augmenting Trees

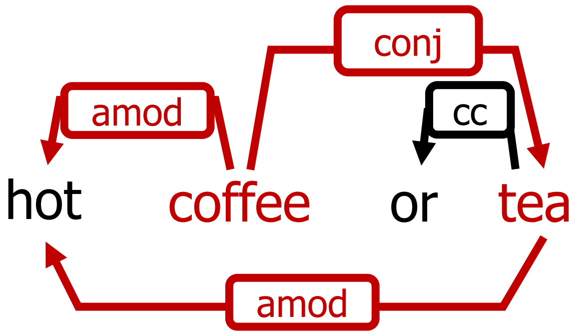


Shi and Lee (ACL, 2020)

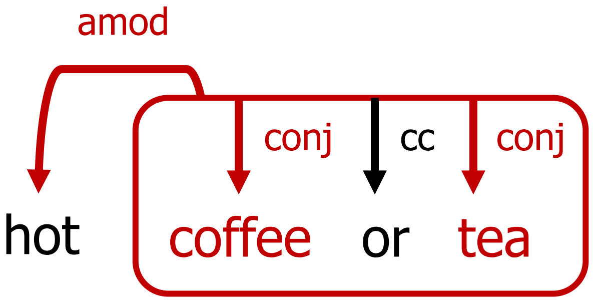


Shi and Lee (EMNLP, 2018)

Beyond Trees



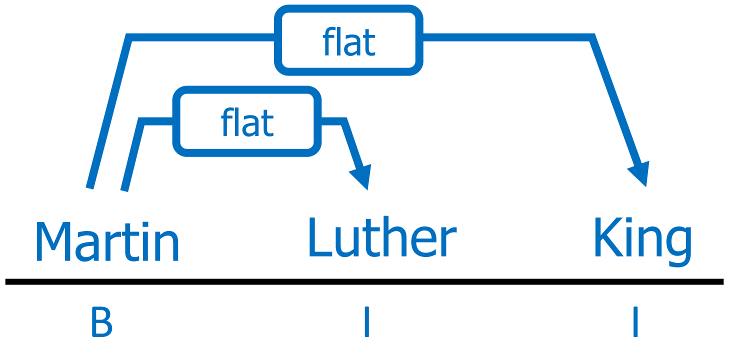
Shi and Lee (IWPT, 2021)



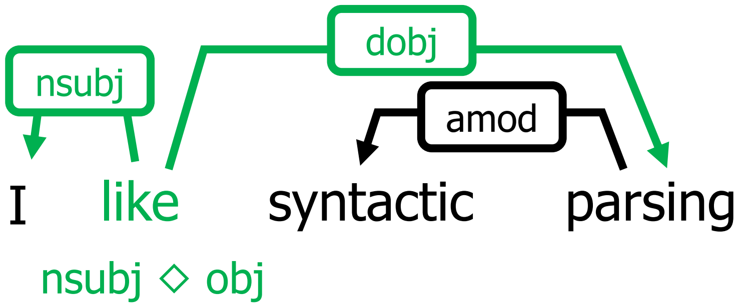
Shi and Lee (ACL, 2021)

Outline

Augmenting Trees

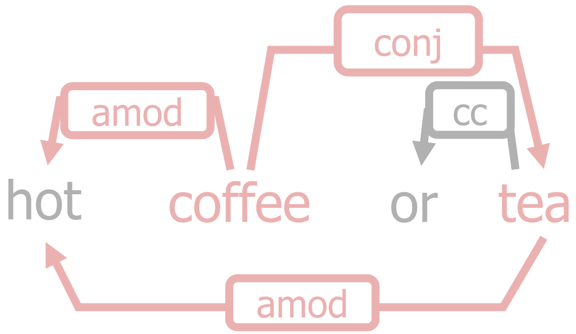


Shi and Lee (ACL, 2020)

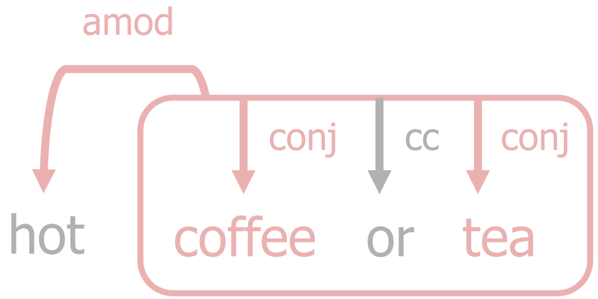


Shi and Lee (EMNLP, 2018)

Beyond Trees



Shi and Lee (IWPT, 2021)



Shi and Lee (ACL, 2021)

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

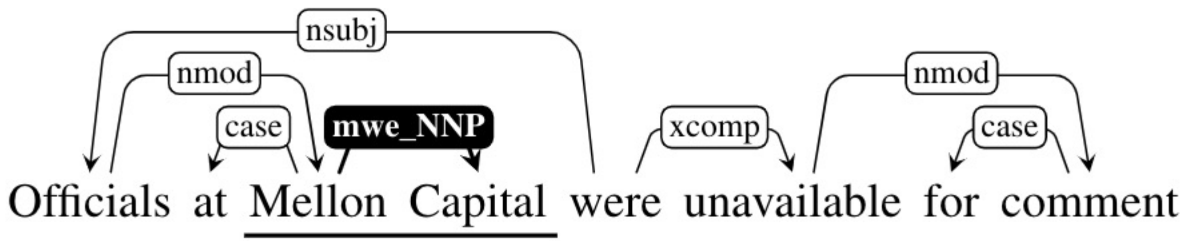
- BIO tagging is a common solution for span extraction, e.g., NER

A	monument	to	Martin	Luther	King
O	O	O	B	I	I

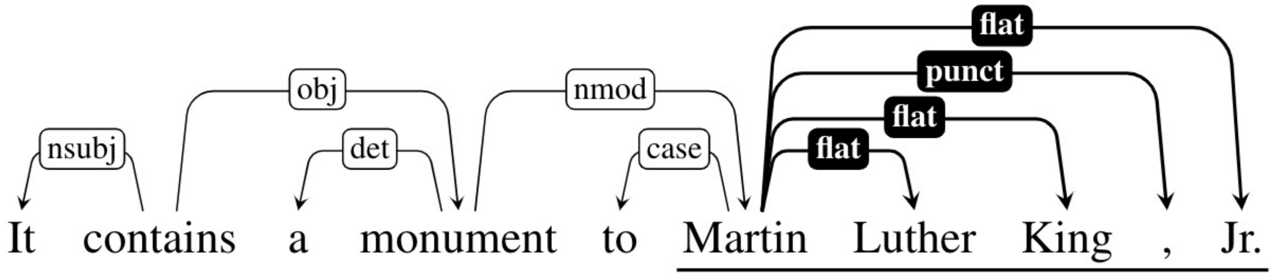
Headless MWEs in Treebanks

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

(Universal Dependencies annotation guideline)

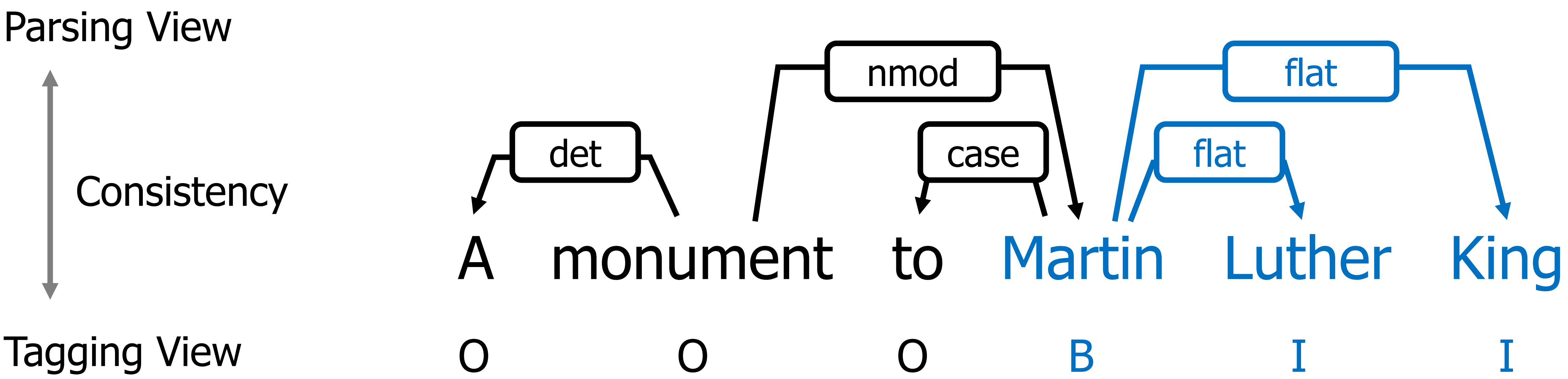


(The MWE-Aware English Dependency Corpus)



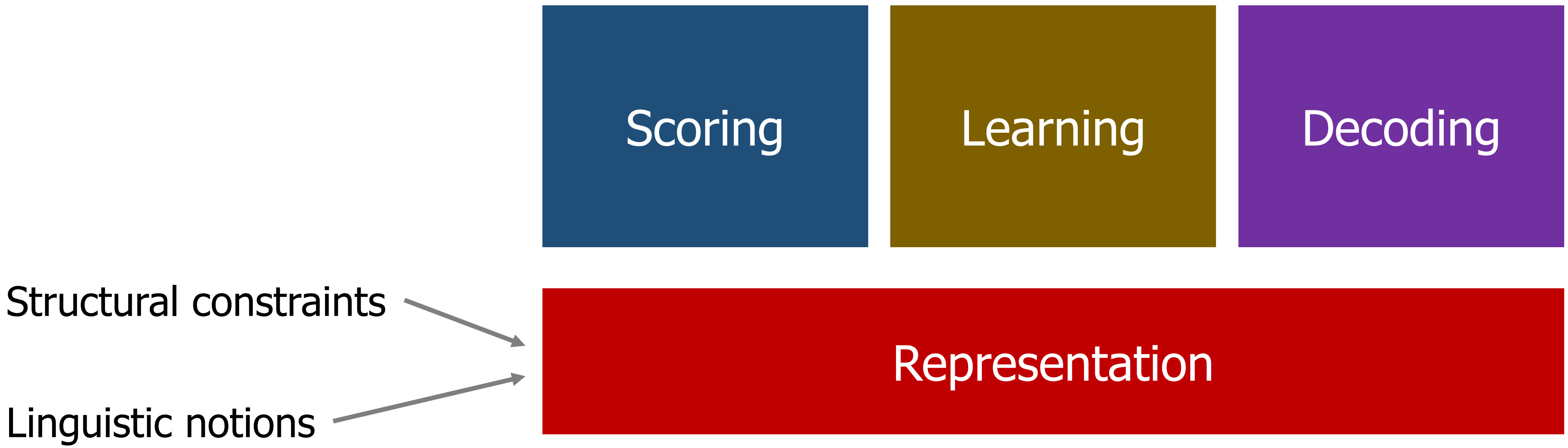
(Universal Dependencies)

Main Idea



Parsing View
Consistency
Tagging View
nmod
flat
det
case
flat
A monument to Martin Luther King
O O O B I I

This Dissertation



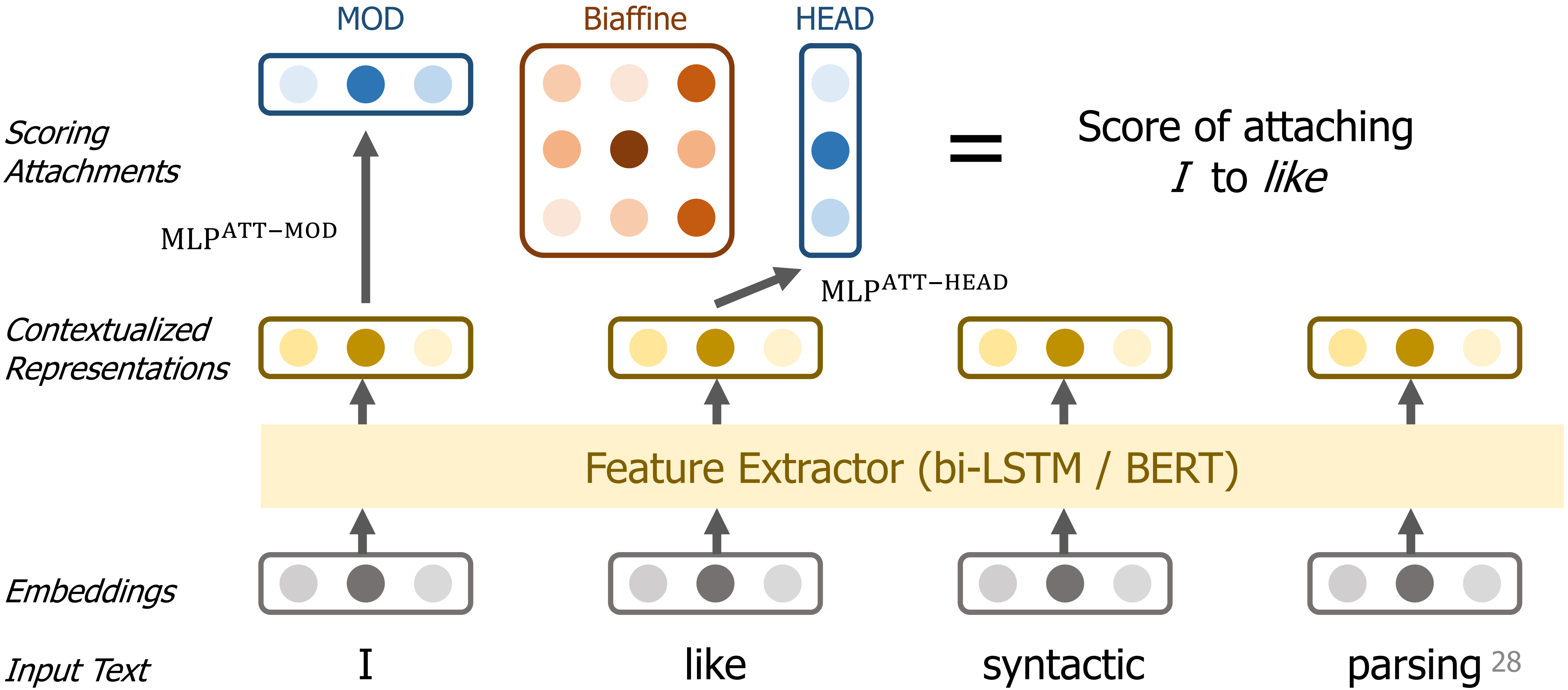
Scoring

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

$$P(y|x) = \frac{1}{Z_x} \prod_{i=1}^{n} \underbrace{P(h_i|x_i)}_{\text{Attachment}} \underbrace{P(r_i|x_i, h_i)}_{\text{Relation labeling}} \underbrace{P(t_i|x_i)}_{\text{MWE BIO Tagging}}$$

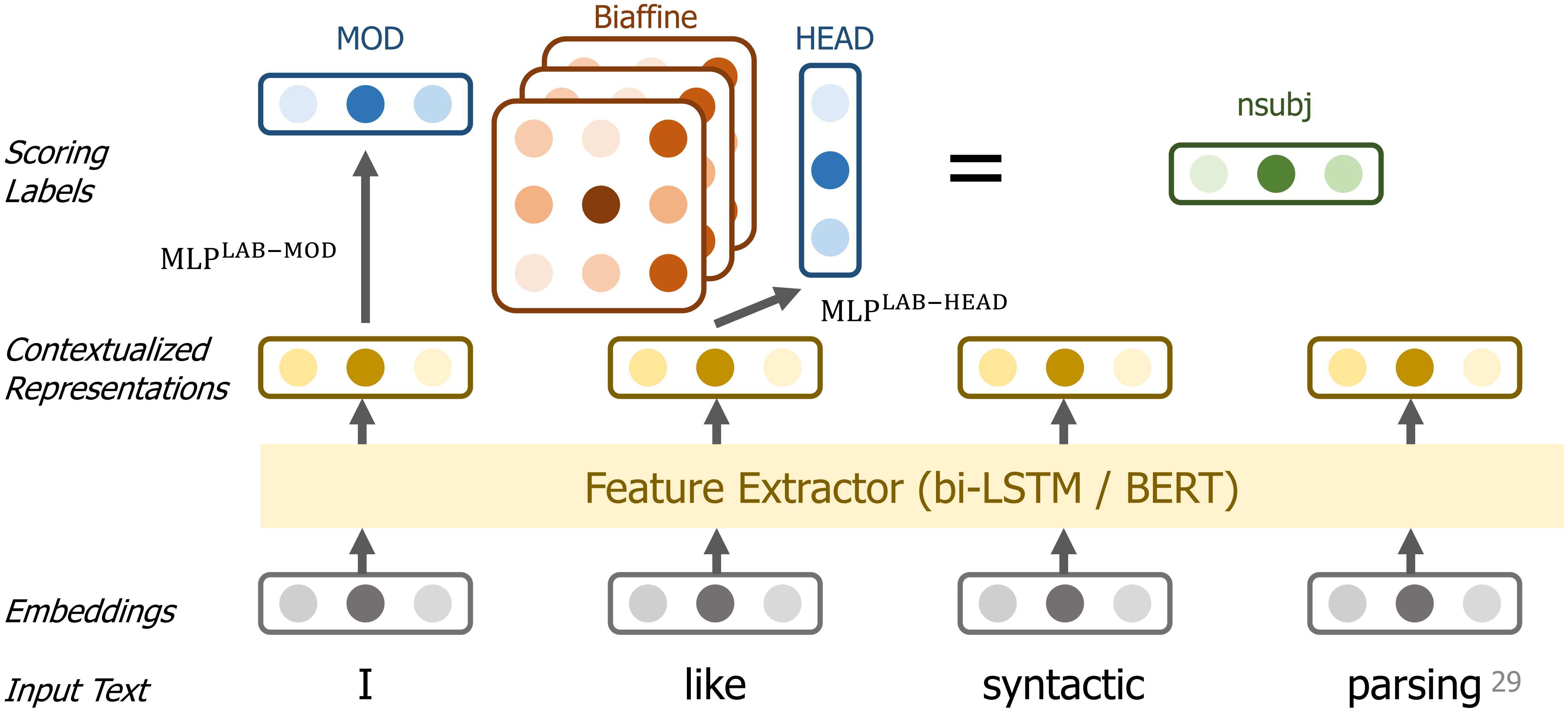
$$P(y|x) = \frac{1}{Z_x} \prod_{i=1}^{n} P(h_i|x_i) P(r_i|x_i, h_i) P(t_i|x_i)$$

Model: Attachment Scoring



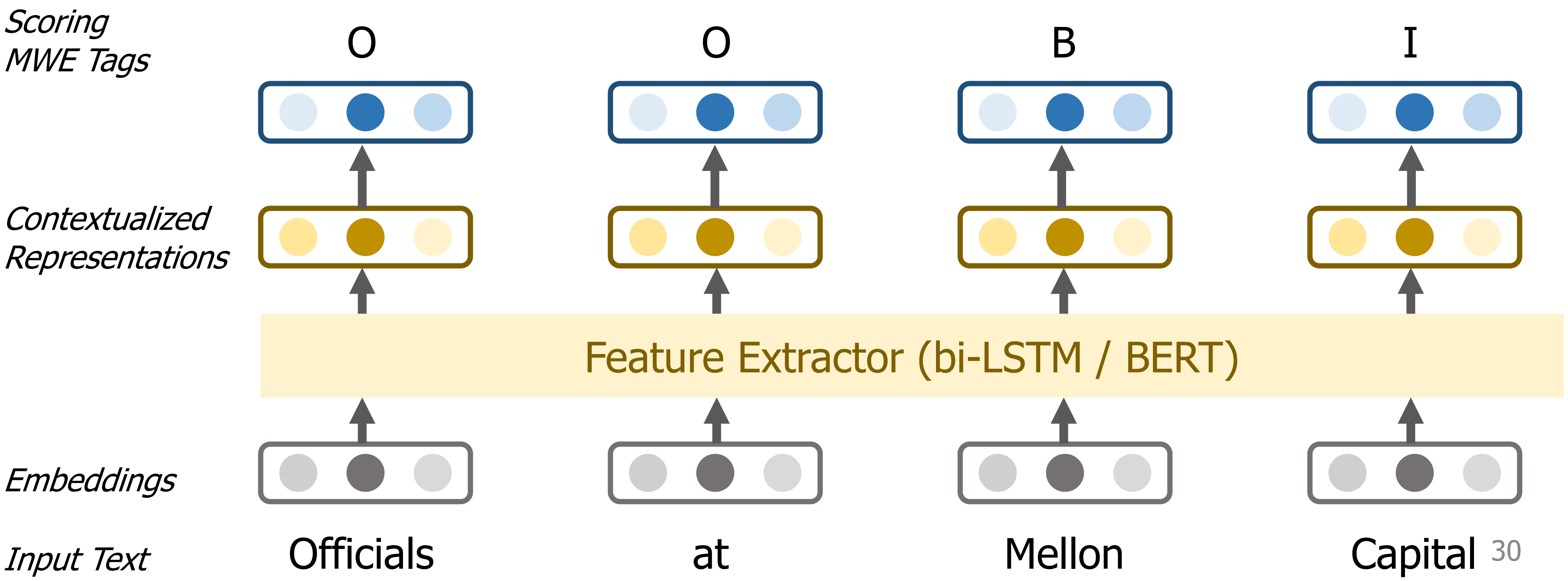
Model: Label Scoring

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

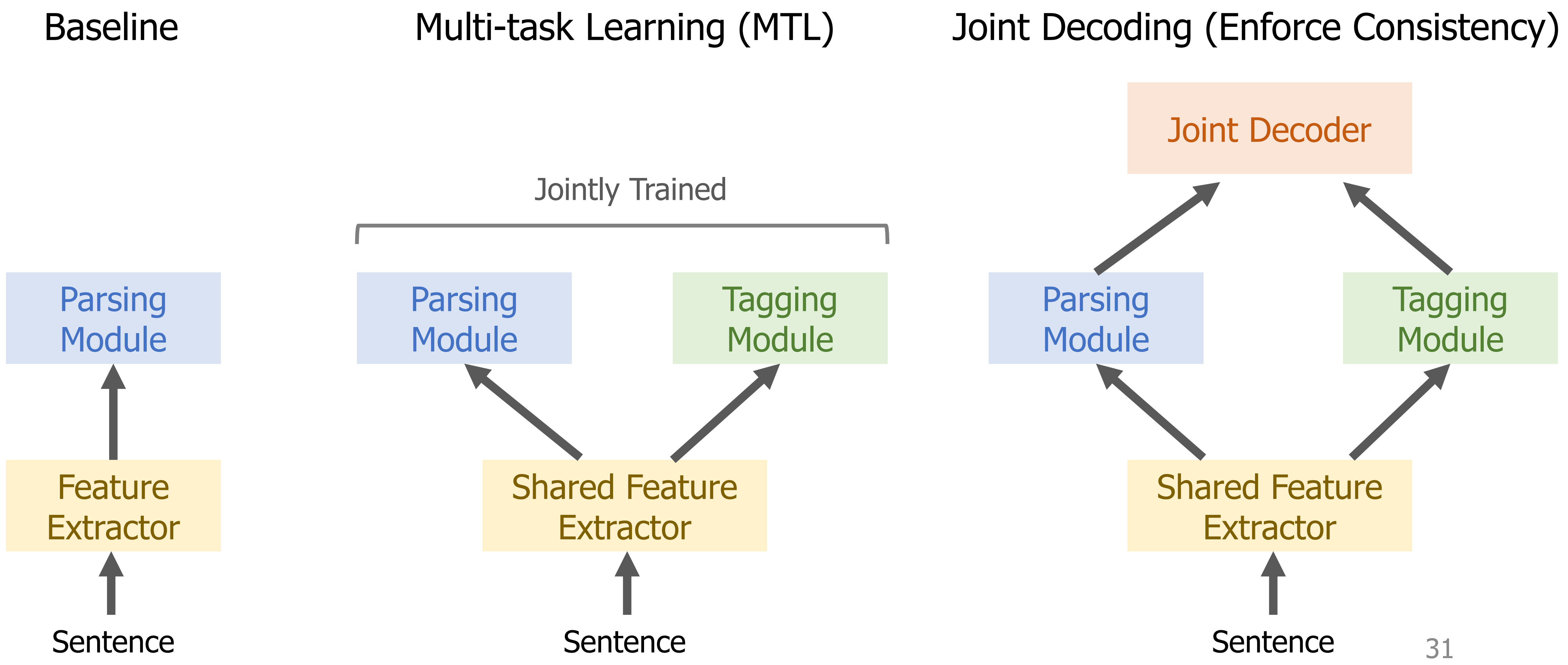


Model: Tagging

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

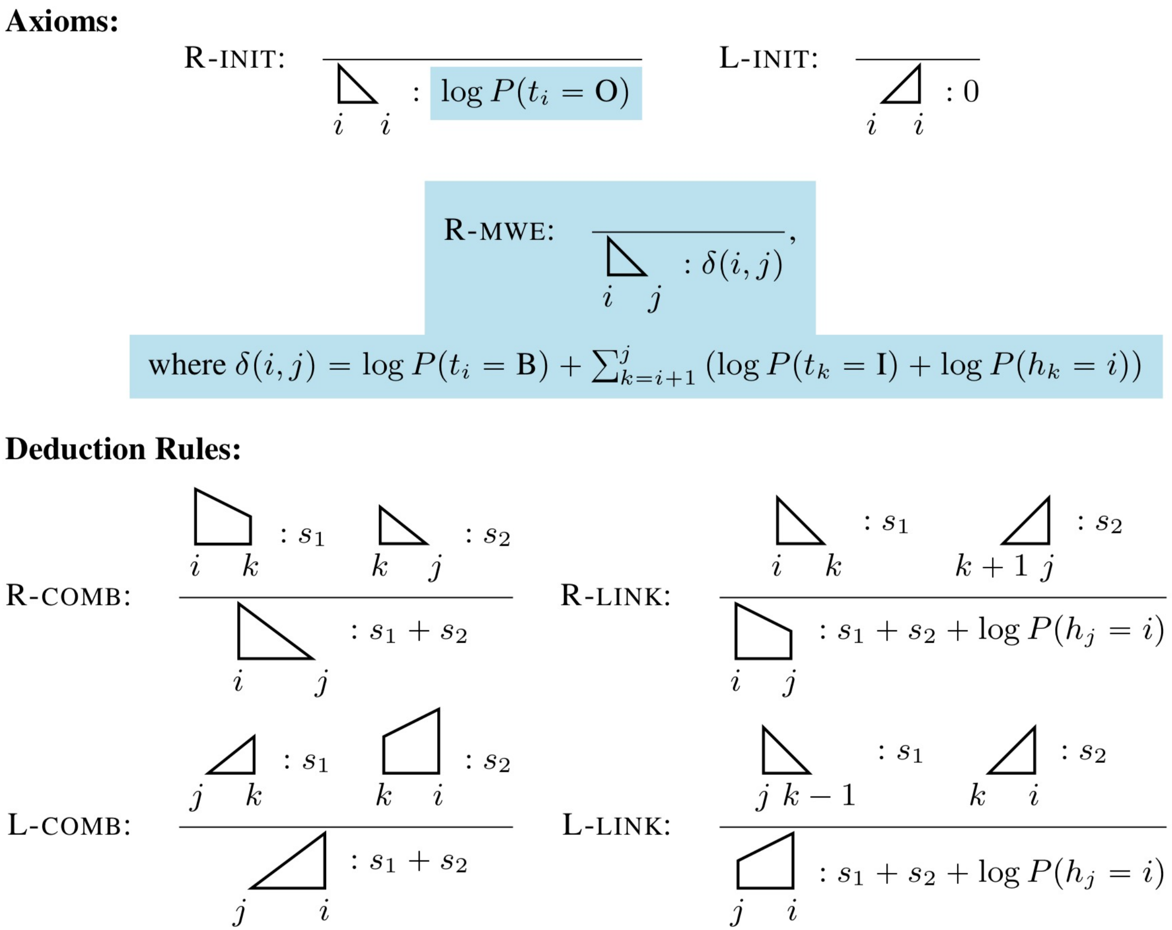


Learning and Inferencing

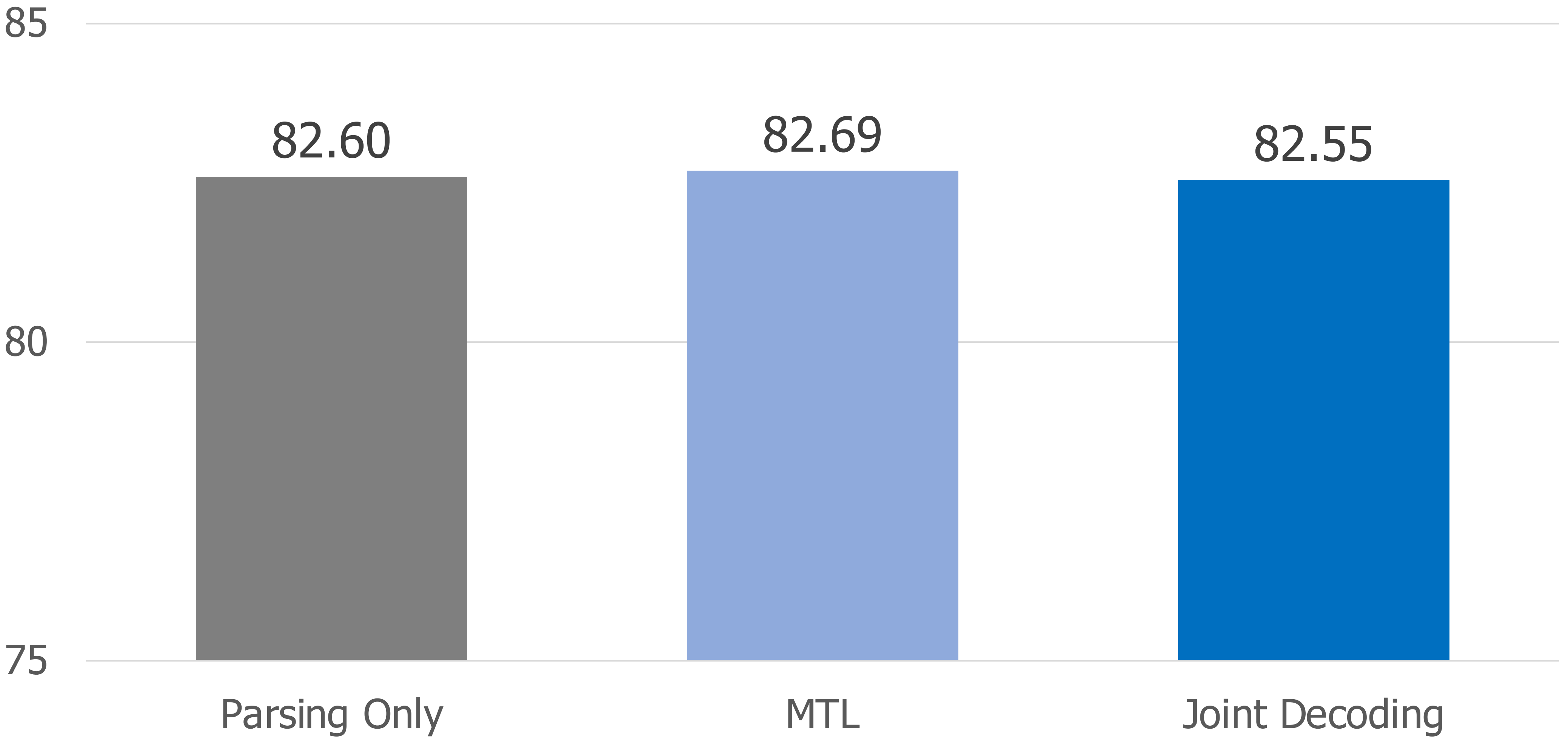


Joint Decoding

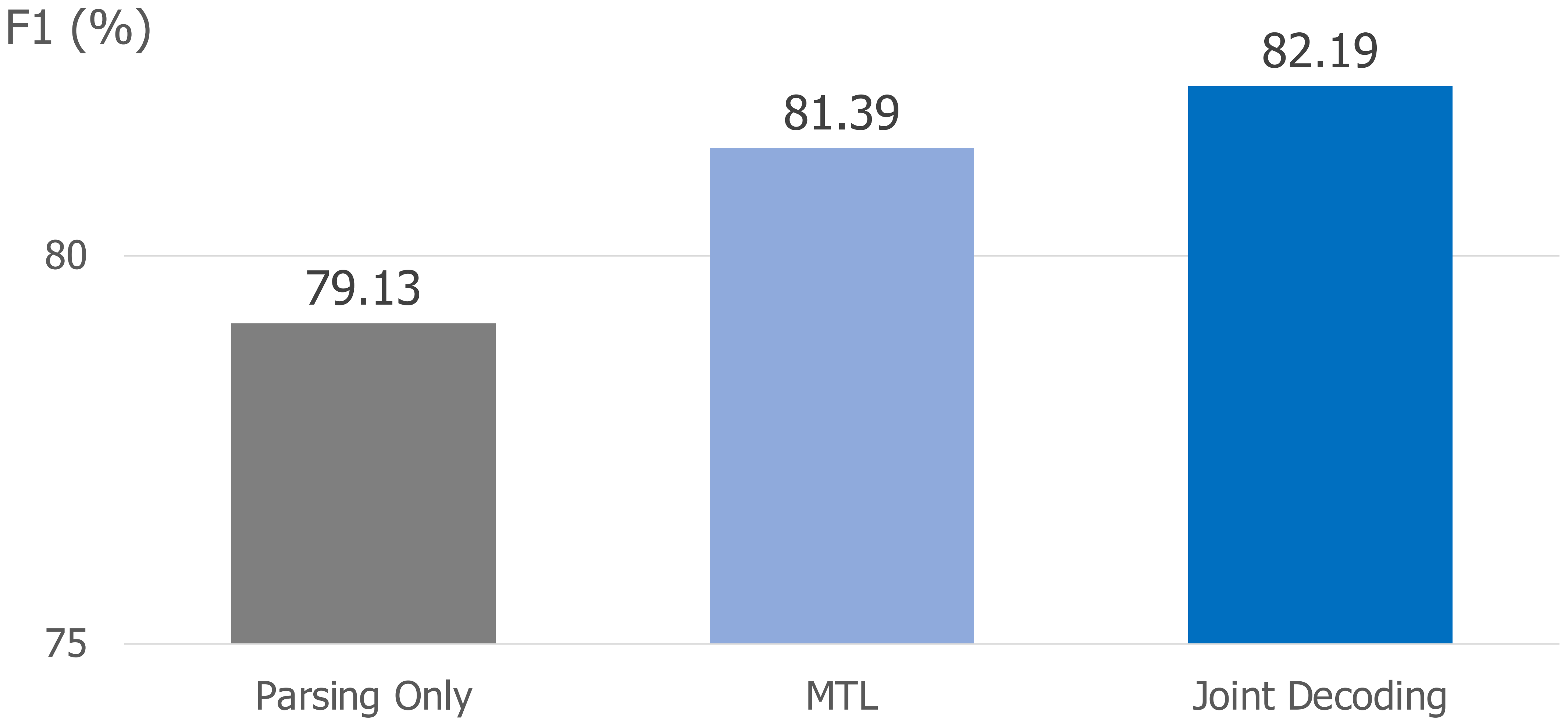
- Key idea: add a deduction rule (axiom) into Eisner's (1996) algorithm



Experiment Results – "Standard" Parsing Metrics



Experiment Results – Headless MWE Identification



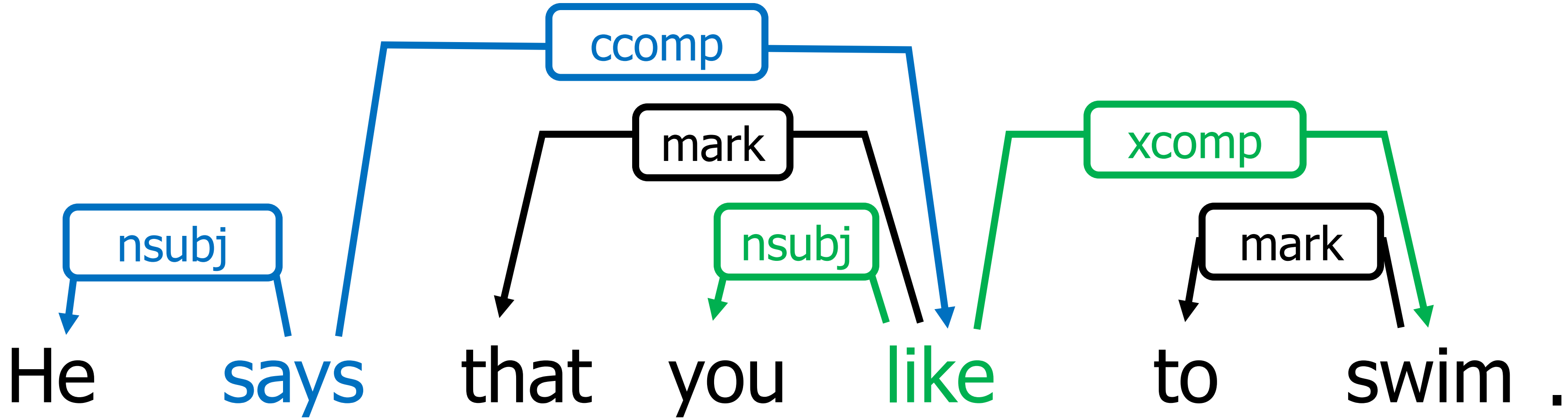
Universal Dependencies Taxonomy

	Nominals	Clauses	Modifier words	Function words
Core arguments	nsubj, obj, iobj	csubj, ccomp, xcomp		
Non-core dependents	obl, vocative, expl, dislocated	advcl	advmod, discourse	aux, cop, mark
Nominal dependents	nmod, appos, nummod	acl	amod	det, clf, case
Coordination	**MWE**	**Loose**	**Special**	**Others**
conj, cc	fixed, flat, compound	list, parataxis	orphan, goeswith, reparandum	punct, root, dep

Valency

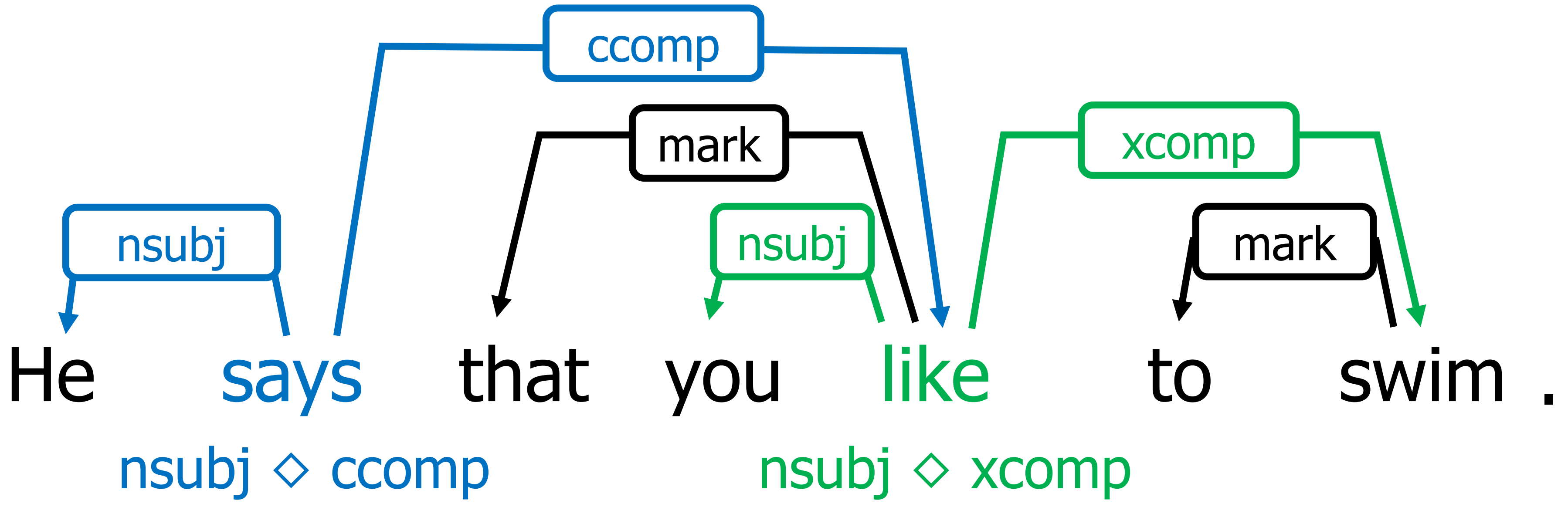
- Valency: Type and number of dependents a word takes

(Tesnière, 1959, *inter alia*)

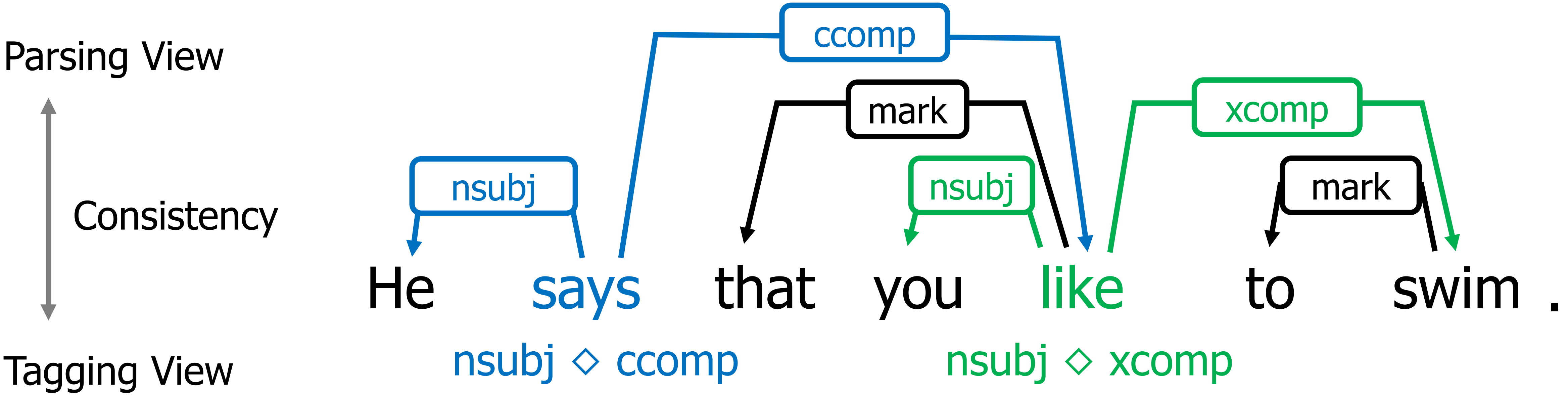


An Empirical Definition of Valency Patterns

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



Main Idea to Incorporate Valency Patterns

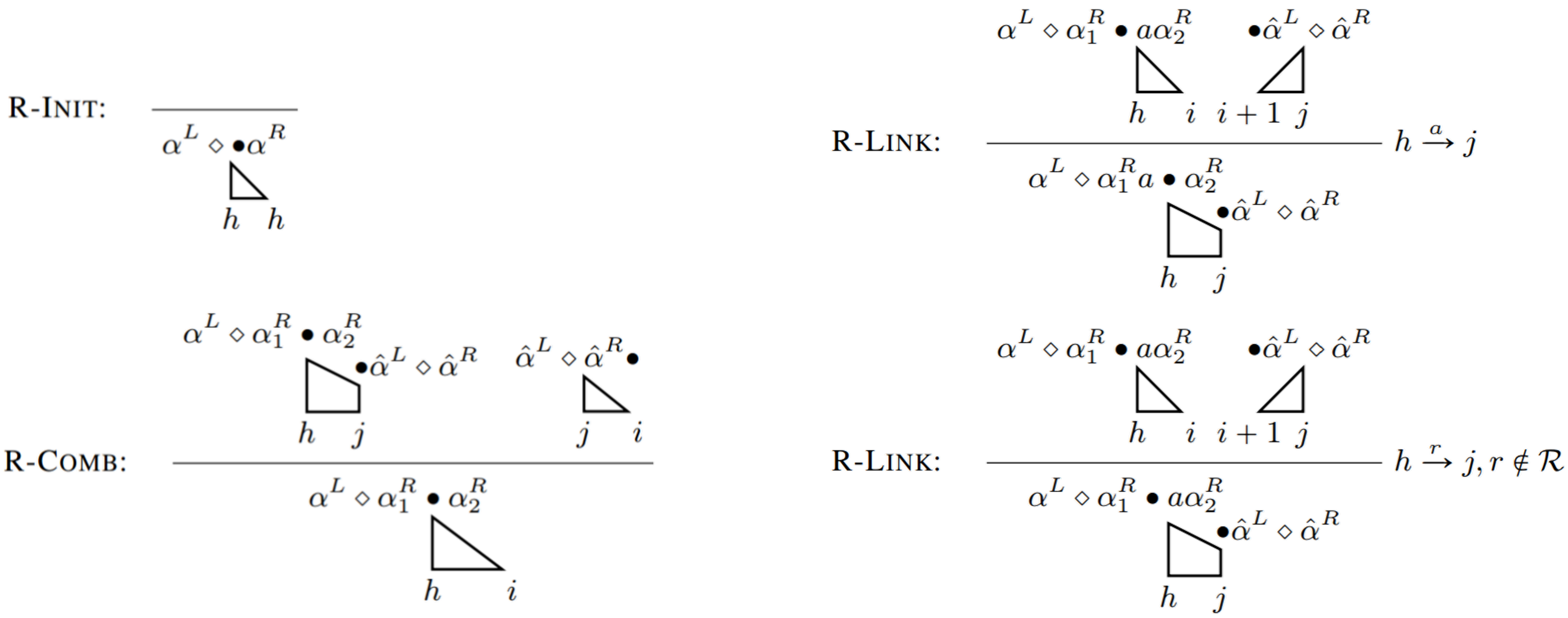


Scoring

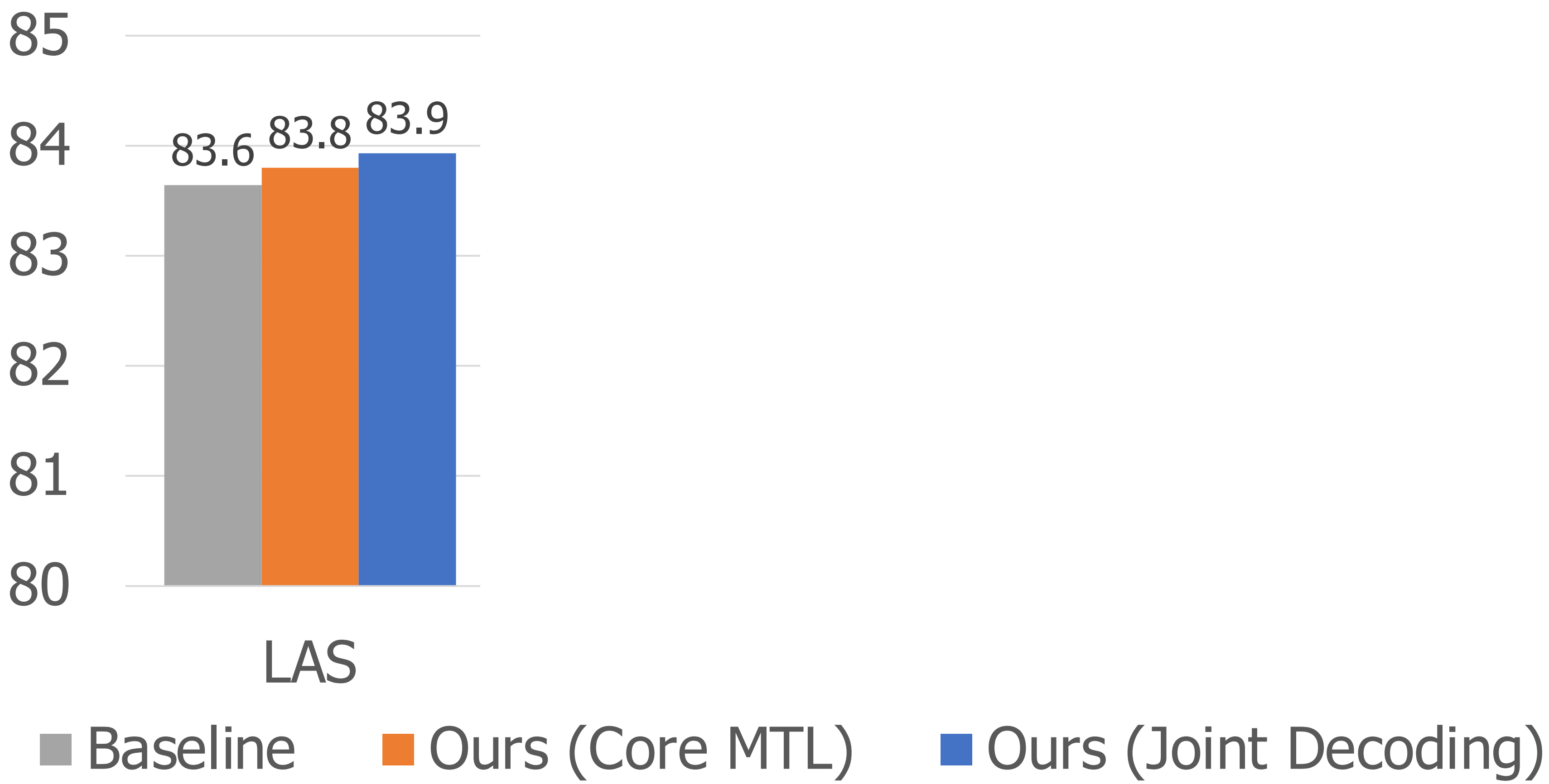
- Dozat and Manning (2017)'s state-of-the-art dependency parser
- + Tagging/Supertagging

$$P(y|x) = \frac{1}{Z_x} \prod_{i=1}^{n} \underbrace{P(h_i|x_i)}_{\text{Attachment}} \underbrace{P(r_i|x_i, h_i)}_{\text{Relation labeling}} \underbrace{P(t_i|x_i)}_{\text{MWE/Valency Tagging}}$$

Decoding with Head Automata (Eisner and Satta, 1999)

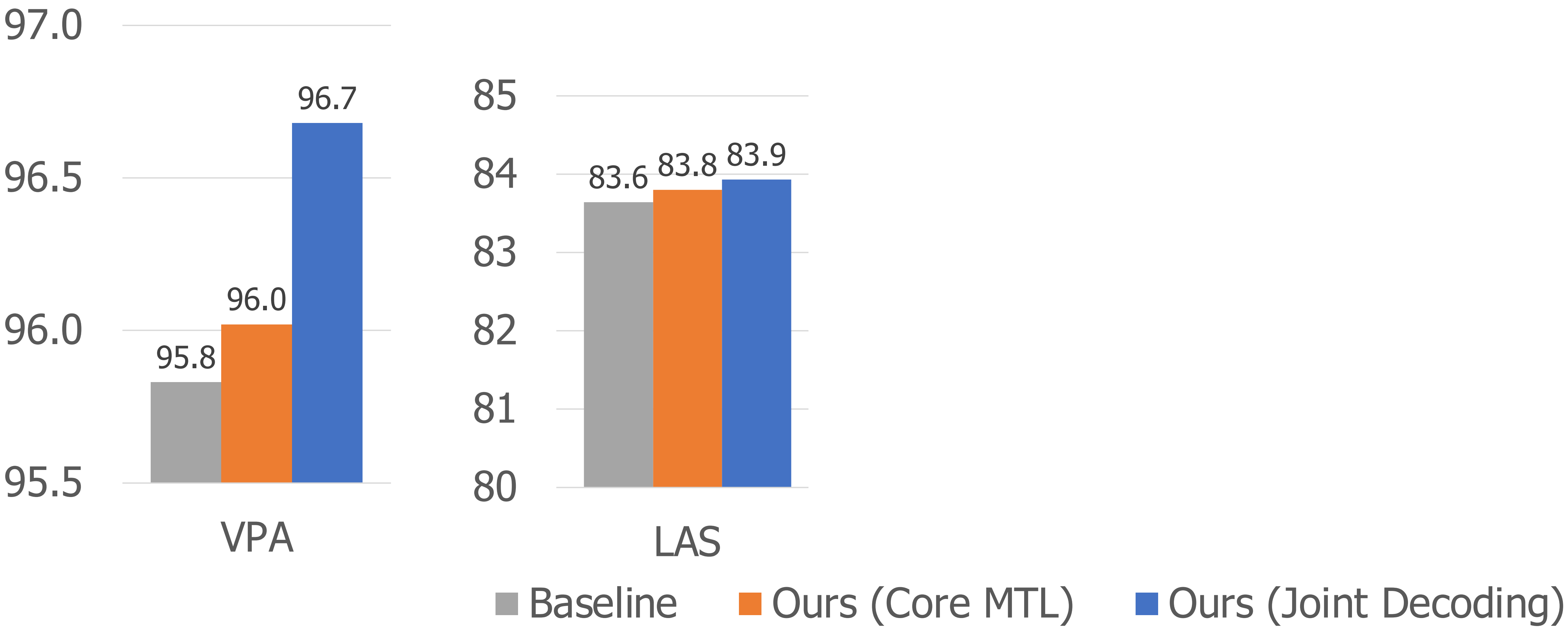


Experiment Results – Valency Augmented Parsing



MTL = Multi-task learning

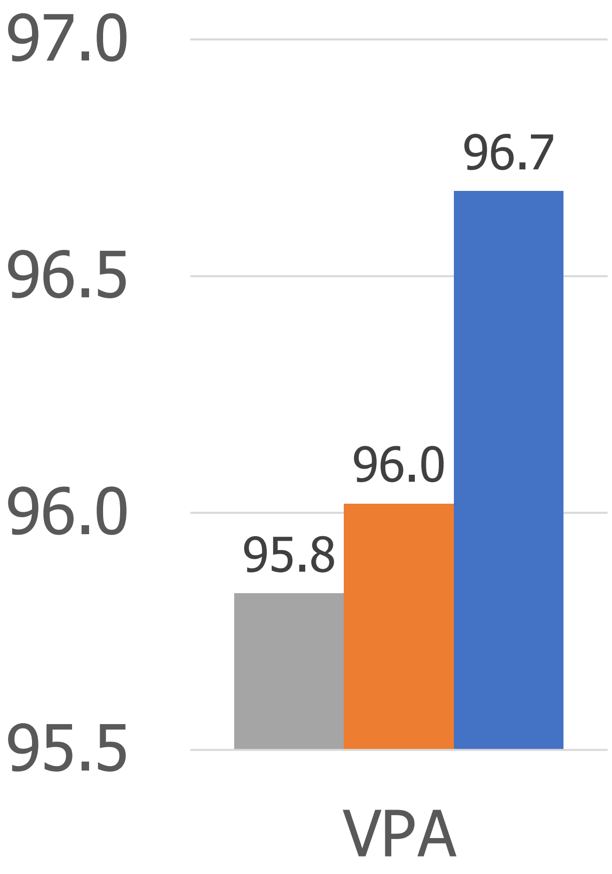
Experiment Results – Valency Augmented Parsing

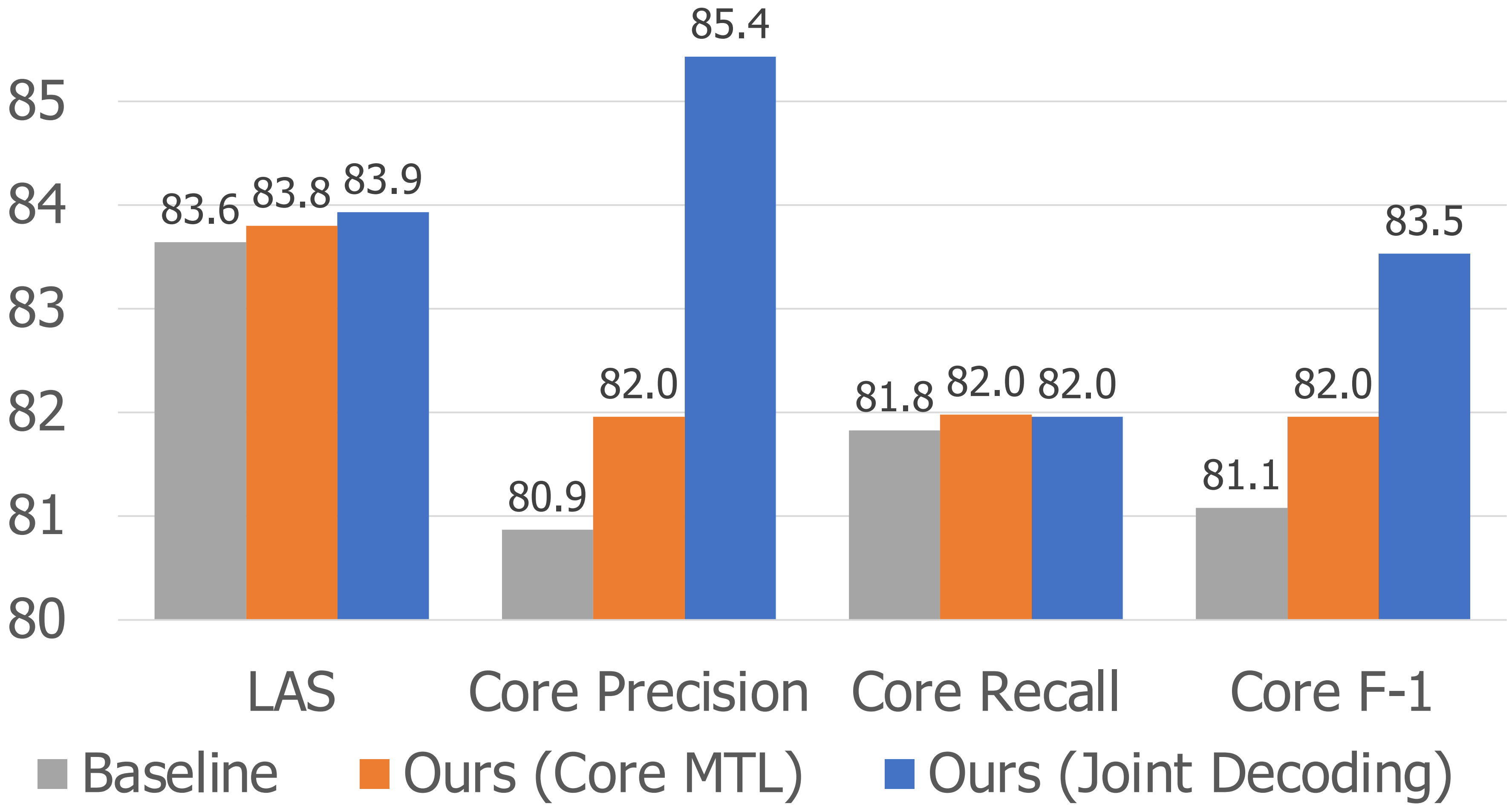


VPA = Valency pattern accuracy

MTL = Multi-task learning

Experiment Results – Valency Augmented Parsing



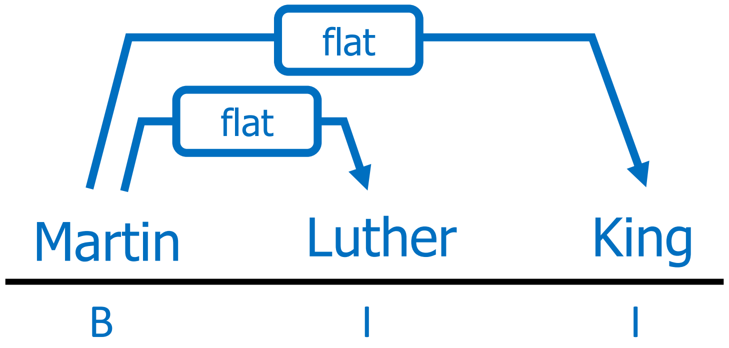


VPA = Valency pattern accuracy

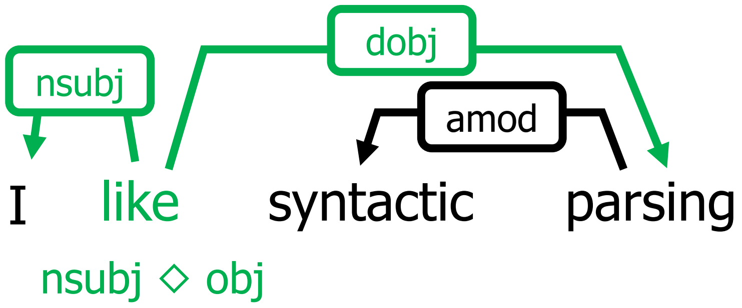
MTL = Multi-task learning

Outline

Augmenting Trees

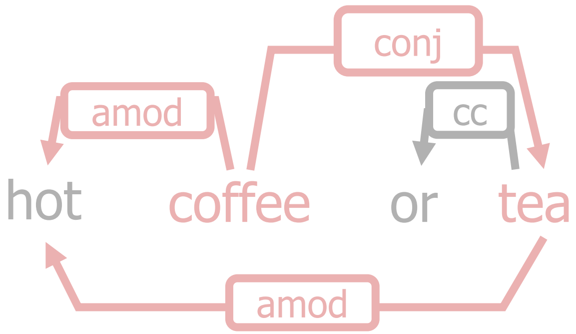


Shi and Lee (ACL, 2020)

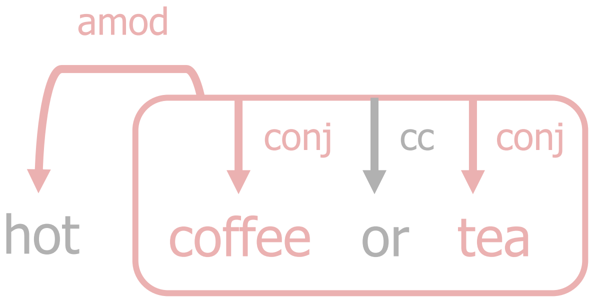


Shi and Lee (EMNLP, 2018)

Beyond Trees



Shi and Lee (IWPT, 2021)

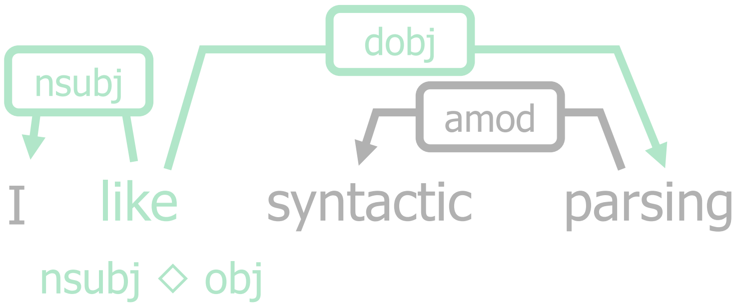


Shi and Lee (ACL, 2021)

Outline

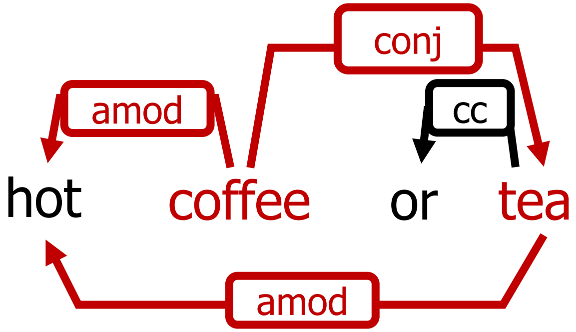
Augmenting Trees

Shi and Lee (ACL, 2020)

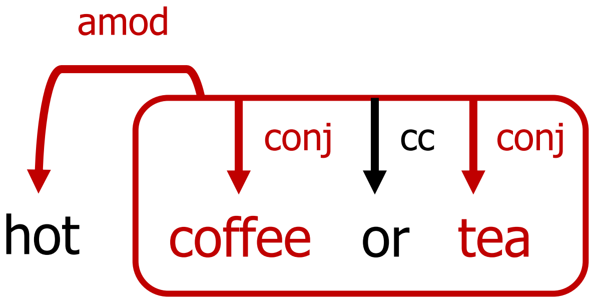


Shi and Lee (EMNLP, 2018)

Beyond Trees

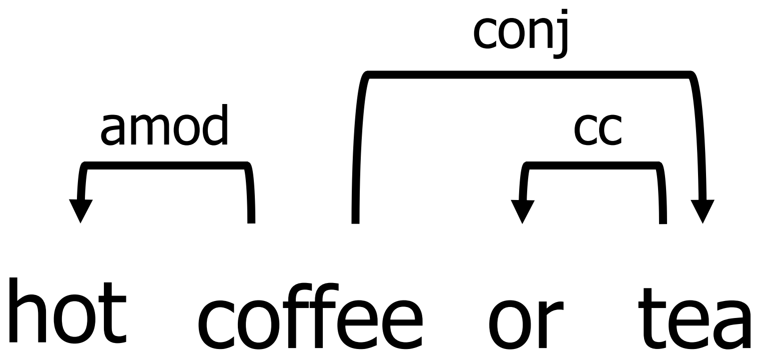


Shi and Lee (IWPT, 2021)

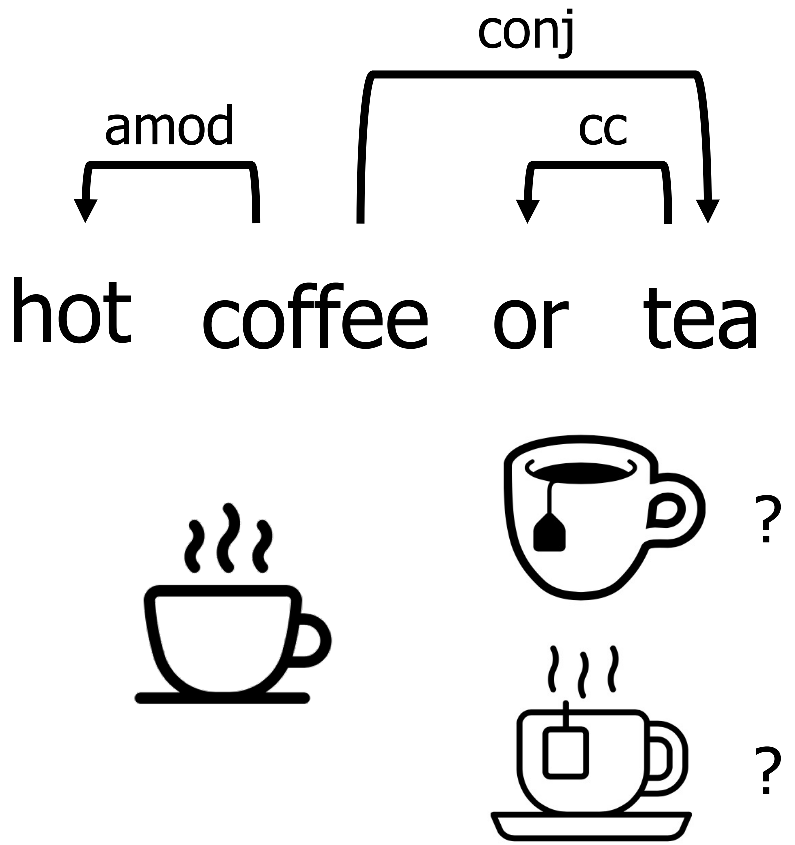


Shi and Lee (ACL, 2021)

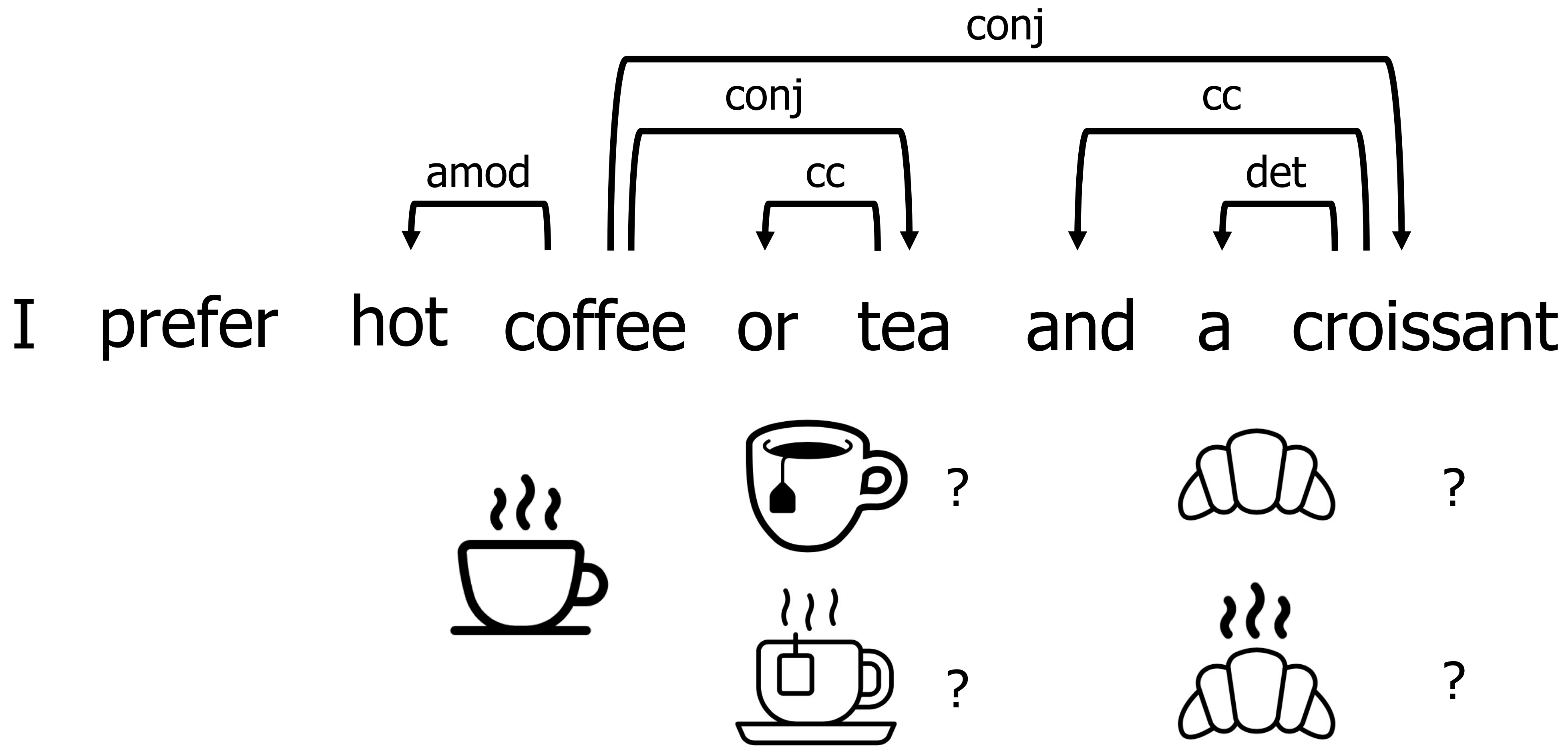
Coordination in Dependency Structures



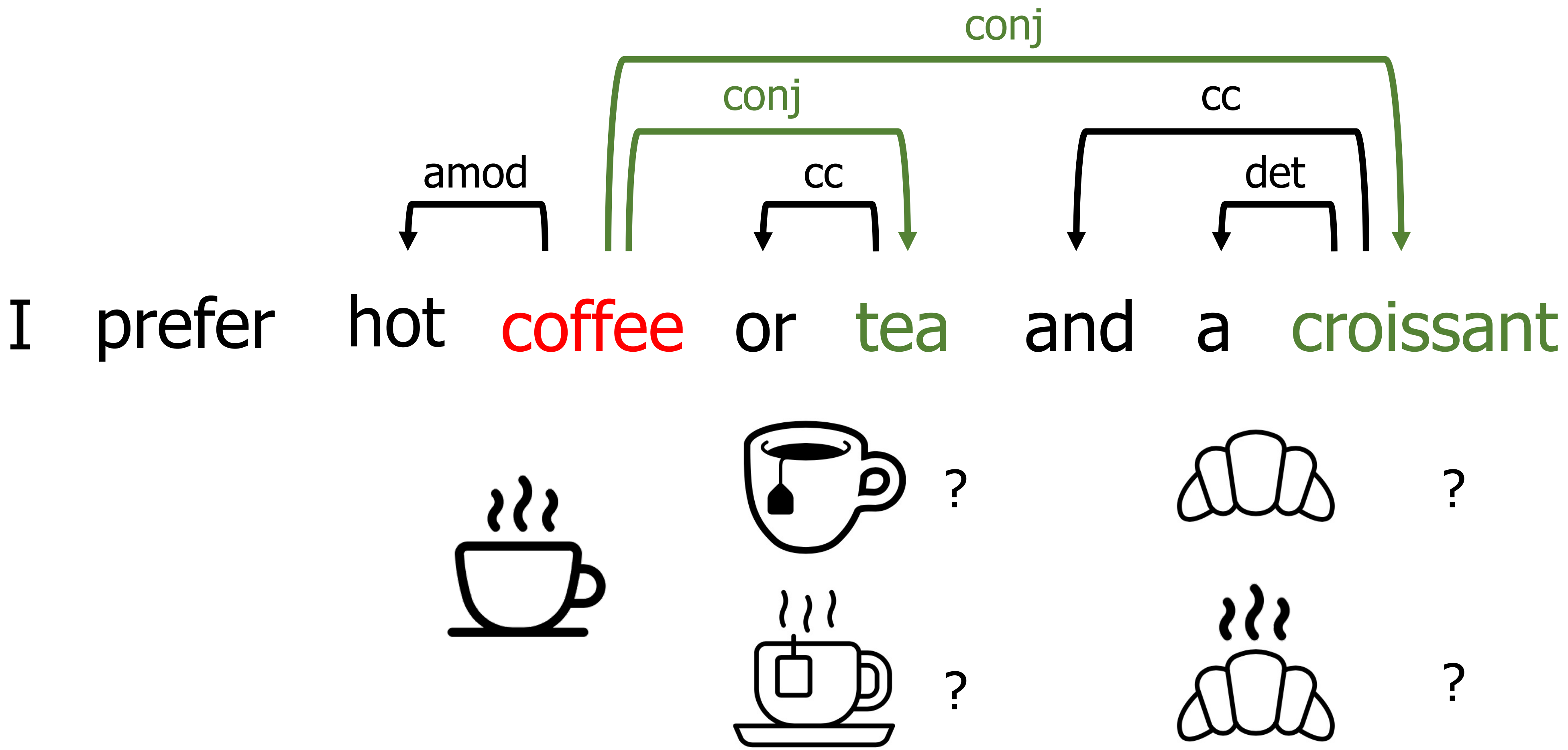
Coordination in Dependency Structures



Coordination in Dependency Structures



Coordination in Dependency Structures



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

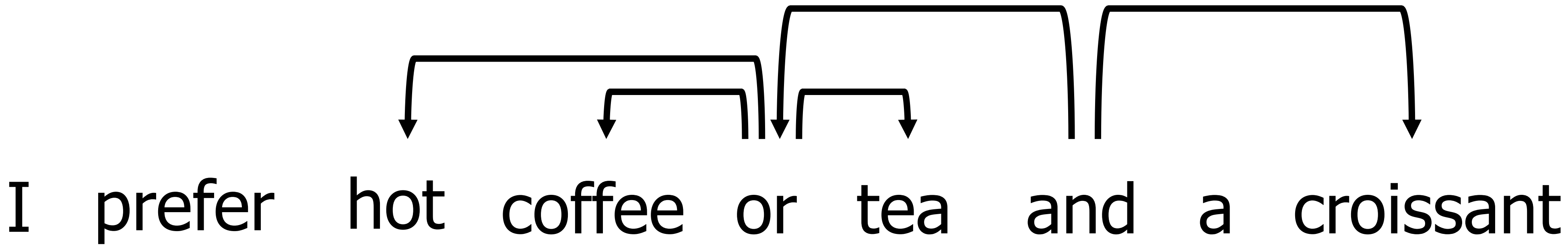
Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 517–527,
Sofia, Bulgaria, August 4-9 2013.

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]

[1]We use the term *modifier* (or *child*) for all types of dependent nodes including *arguments*.

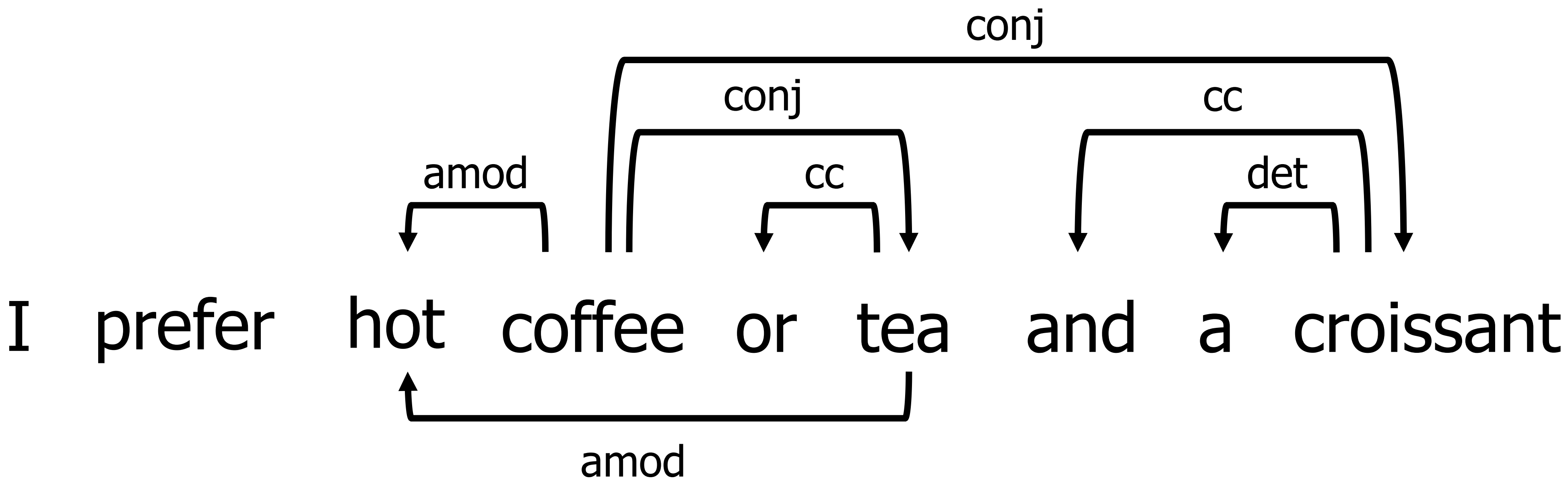
Dependency-based Solutions

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



Dependency-based Solutions

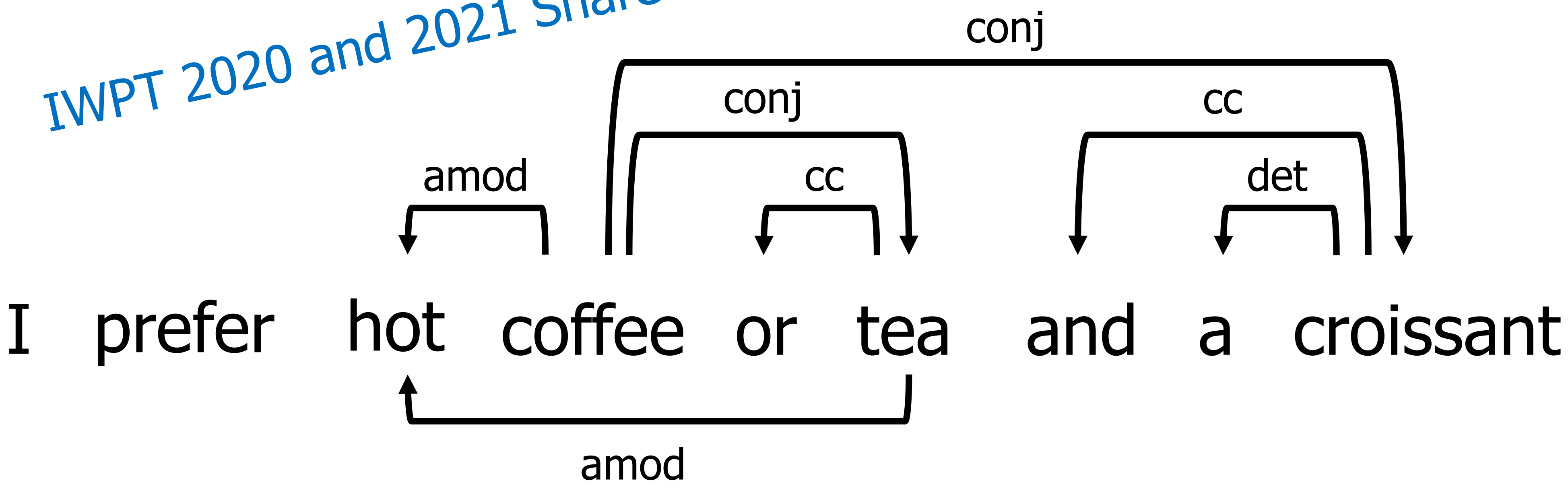
- Enhanced UD Graphs (Schuster and Manning, 2016; Nivre et al., 2018; Bouma et al., 2020)



Dependency-based Solutions

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

IWPT 2020 and 2021 Shared Task



IWPT 2021 Shared Task Official Evaluation

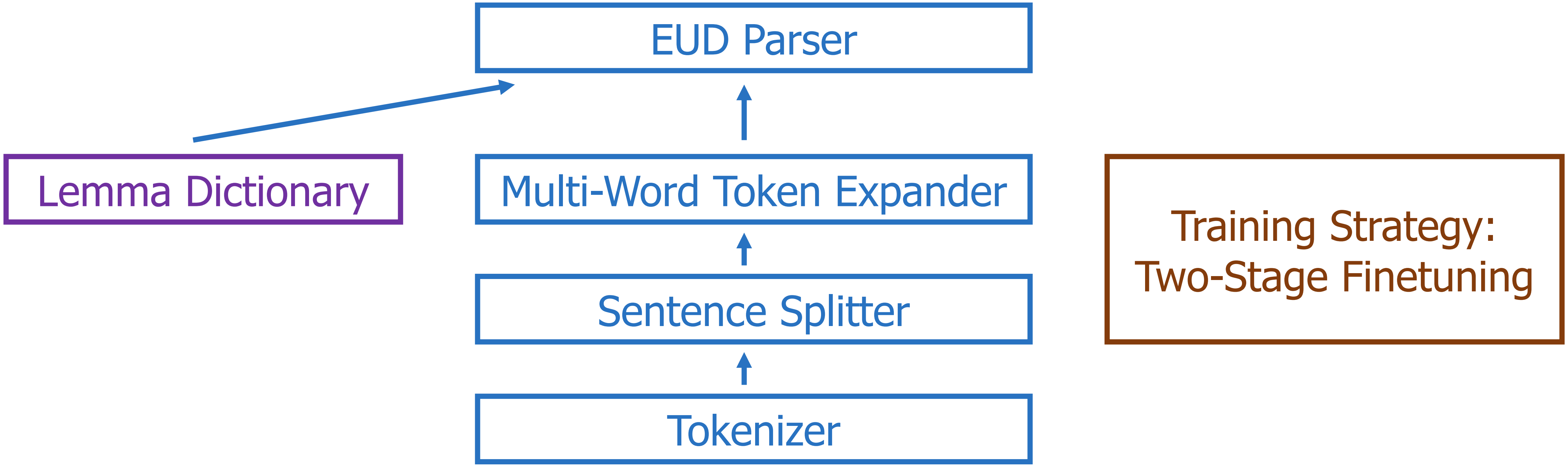
1.	TGIF	89.24
2.	SHANGAITECH	87.07
3.	ROBERTNLP	86.97
4.	COMBO	83.79
5.	UNIPI	83.64
6.	DCU EPFL	83.57
7.	GREW	81.58
8.	FASTPARSE	65.81
9.	NUIG	30.03

2.17 ELAS

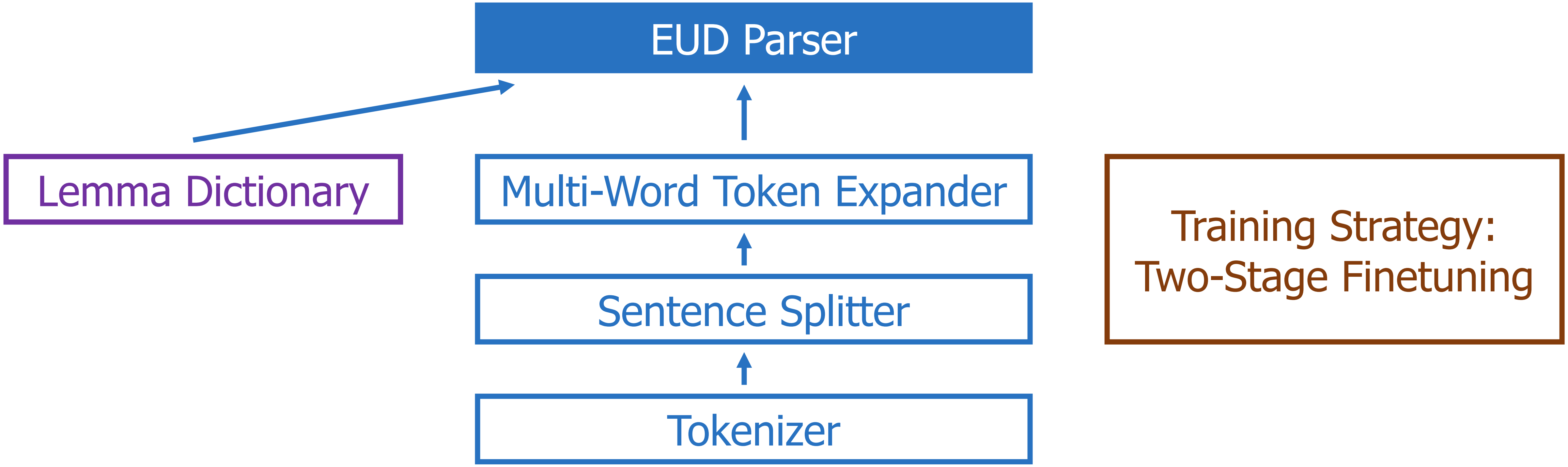
Language	ELAS
Arabic	81.23
Bulgarian	93.63
Czech	92.24
Dutch	91.78
English	88.19
Estonian	88.38
Finnish	91.75
French	91.63
Italian	93.31
Latvian	90.23
Lithuanian	86.06
Polish	91.46
Russian	94.01
Slovak	94.96
Swedish	89.90
Tamil	65.58
Ukrainian	92.78
Average	**89.24**

Best ELAS on 16/17 languages

System Overview

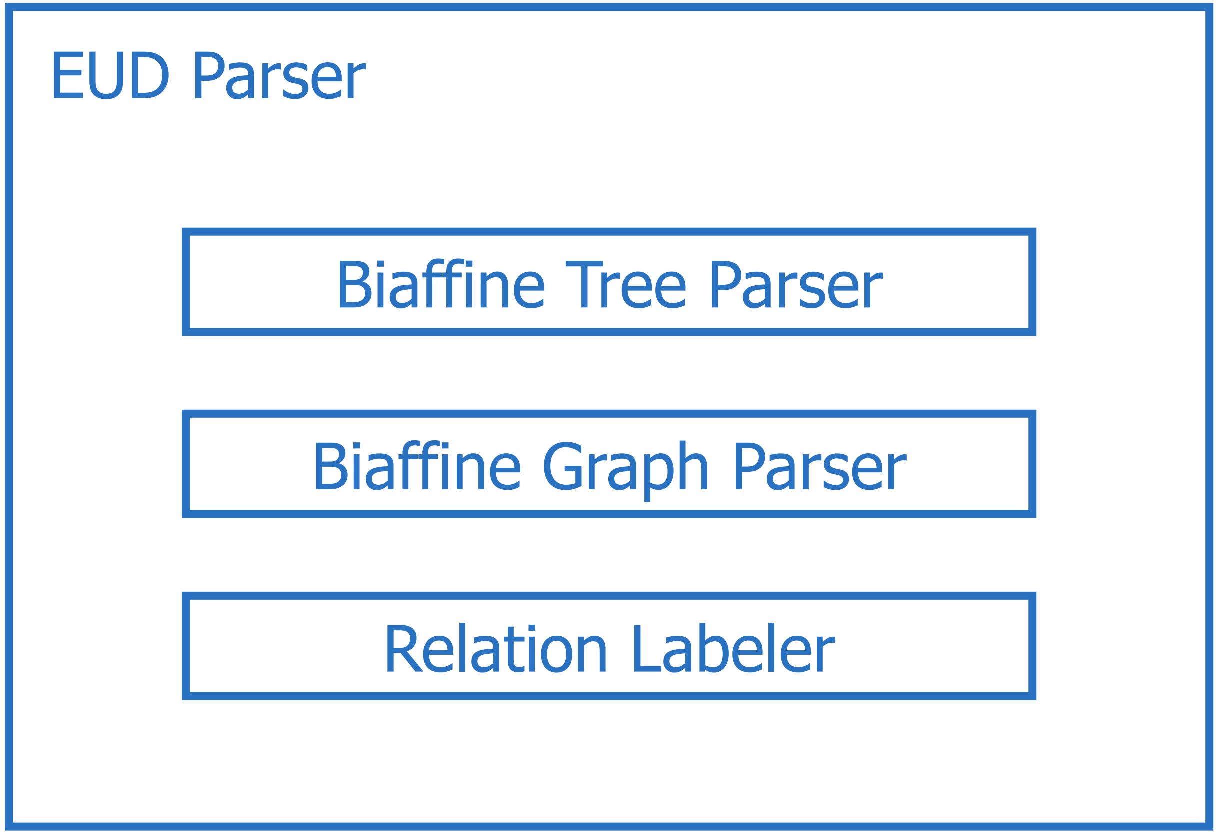


System Overview



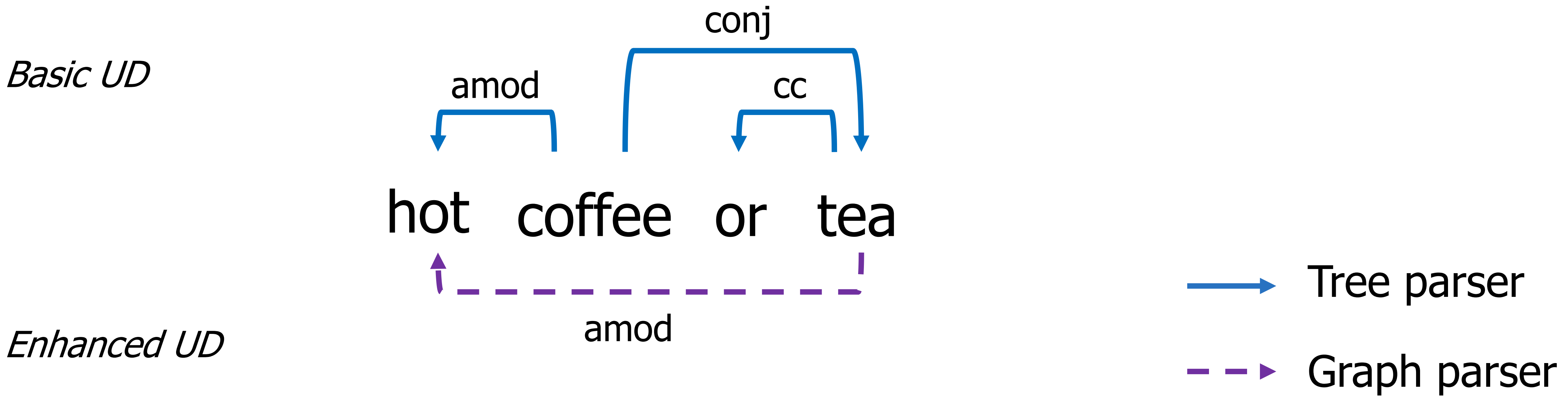
TGIF: **T**ree-**G**raph **I**ntegrated-**F**ormat Parser

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



TGIF: **T**ree-**G**raph **I**ntegrated-**F**ormat Parser

- Every connected graph must have a spanning tree



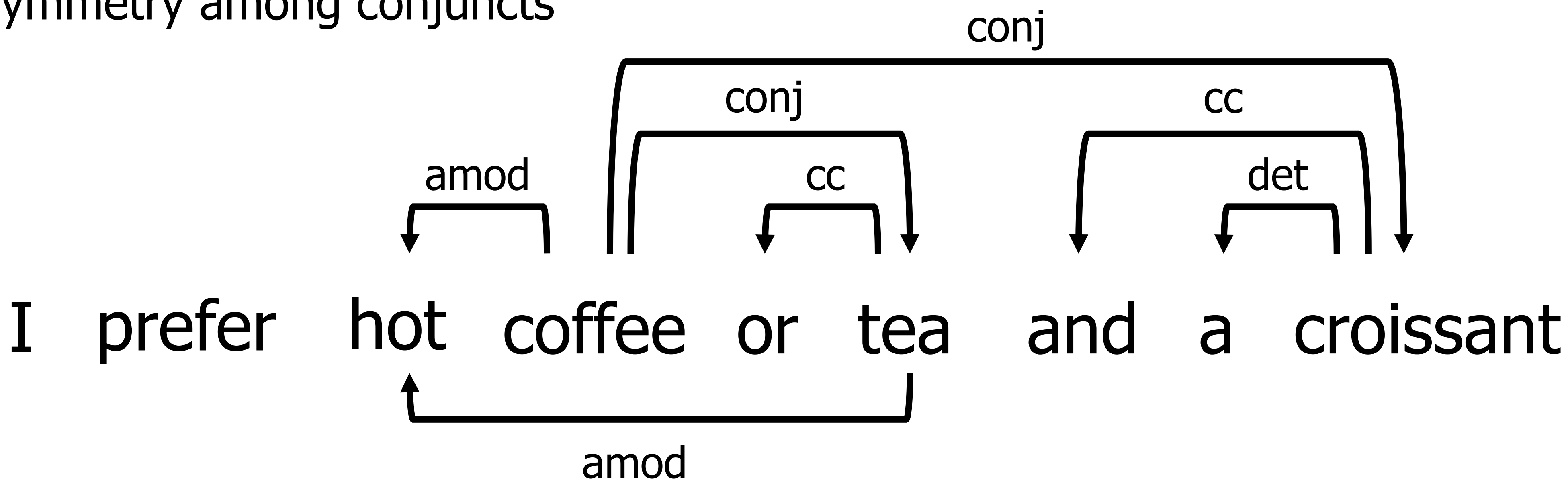
EUD Parsing Results

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

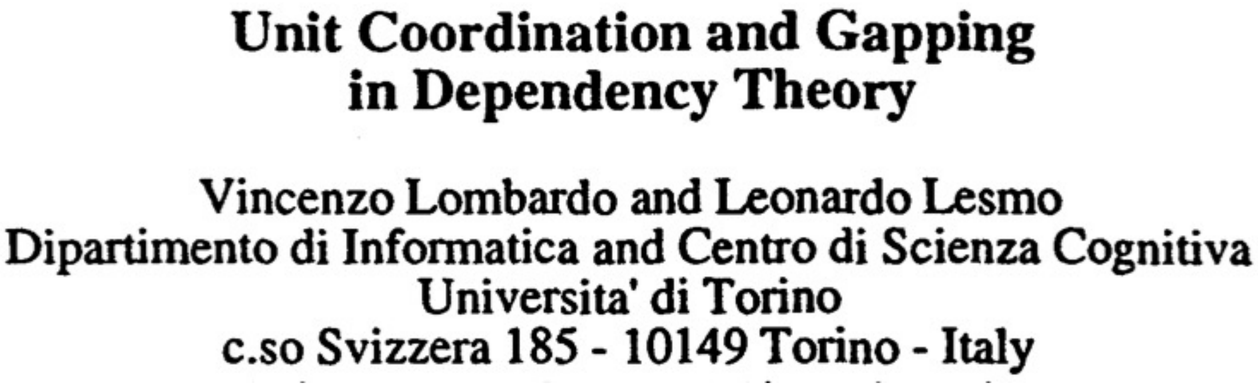
Tree-Graph integrated method wins	Graph-only method wins
Bulgarian, Czech, English, Finnish, French, Italian, Lithuanian, Polish, Russian, Slovak, Swedish, Tamil	Arabic, Dutch, Estonian, Latvian, Ukrainian

EUD Graphs

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



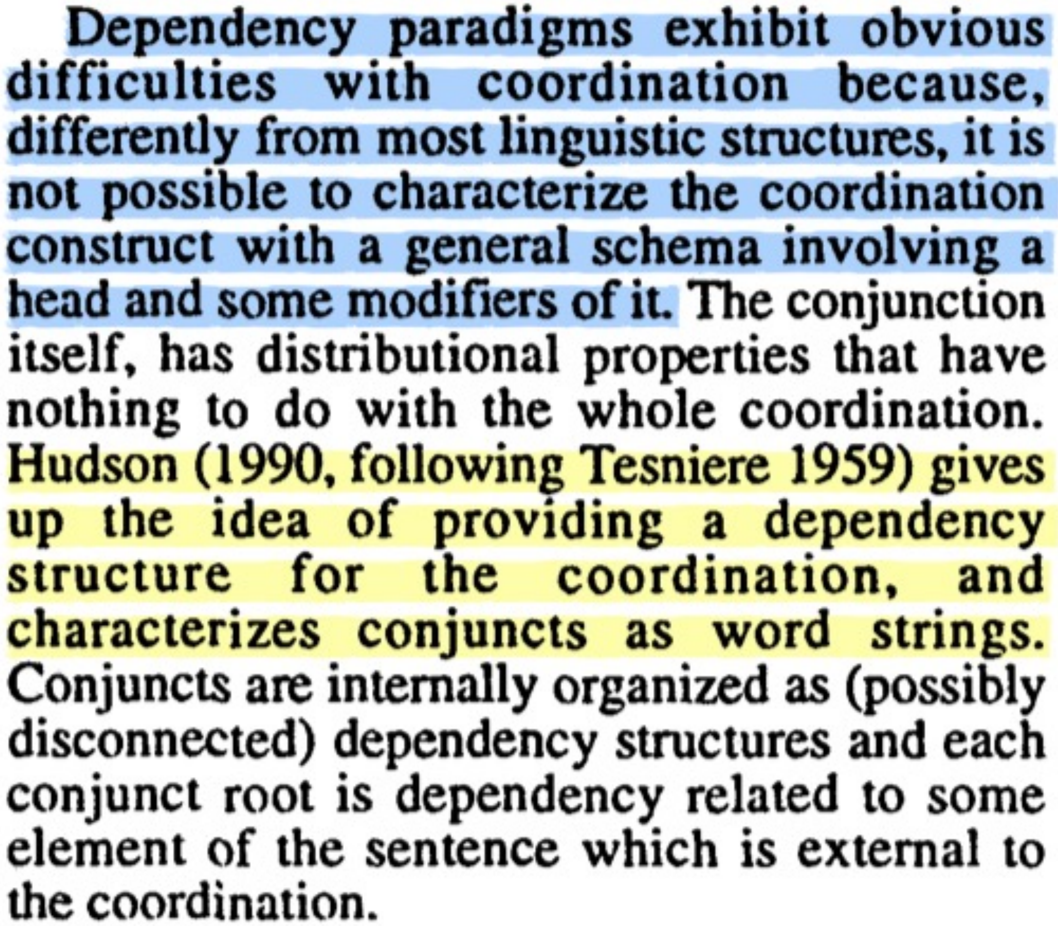
Looking for Other Solutions …



Unit Coordination and Gapping in Dependency Theory

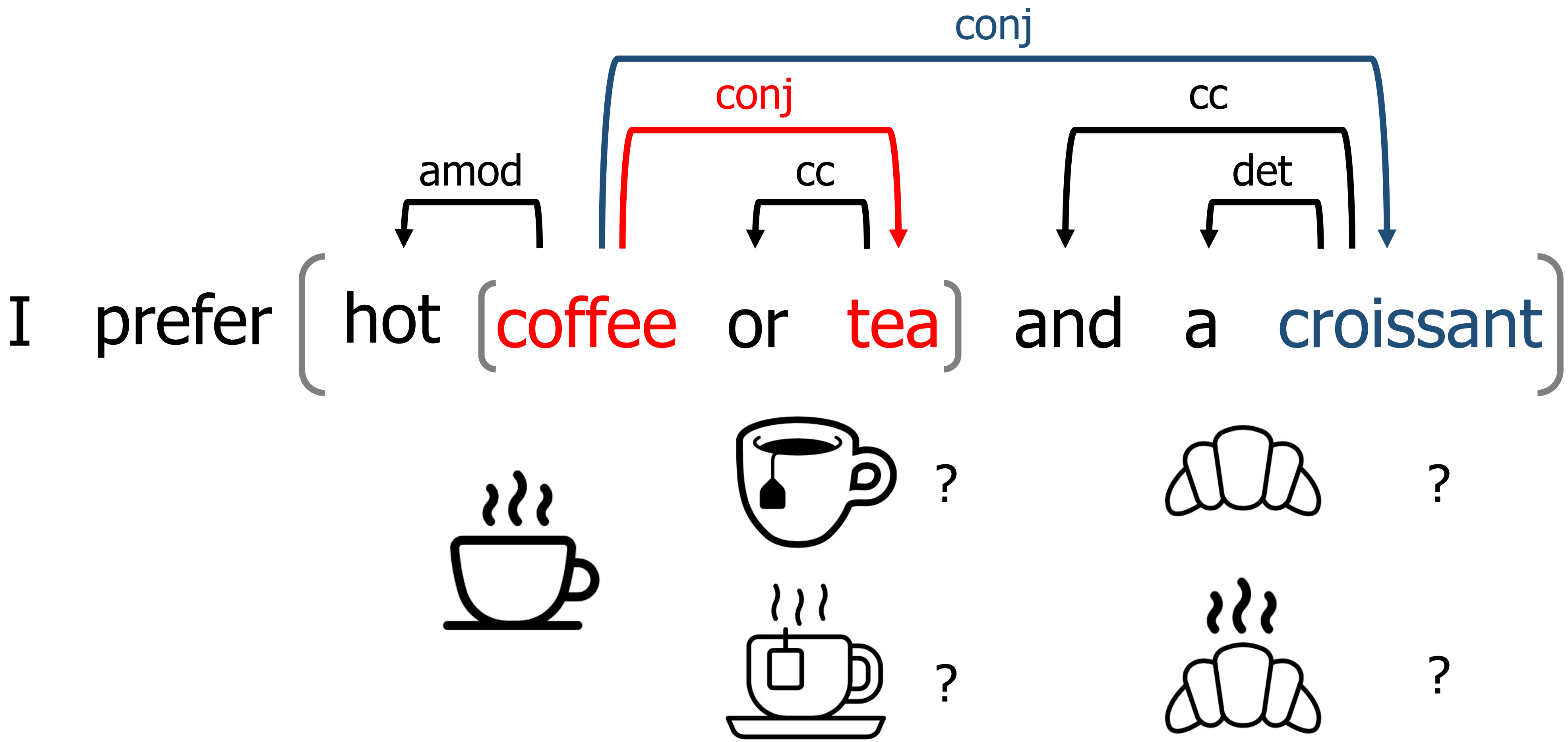
Vincenzo Lombardo and Leonardo Lesmo
Dipartimento di Informatica and Centro di Scienza Cognitiva
Universita' di Torino
c.so Svizzera 185 - 10149 Torino - Italy

Processing of Dependency-Based Grammars (Workshop, 1998)

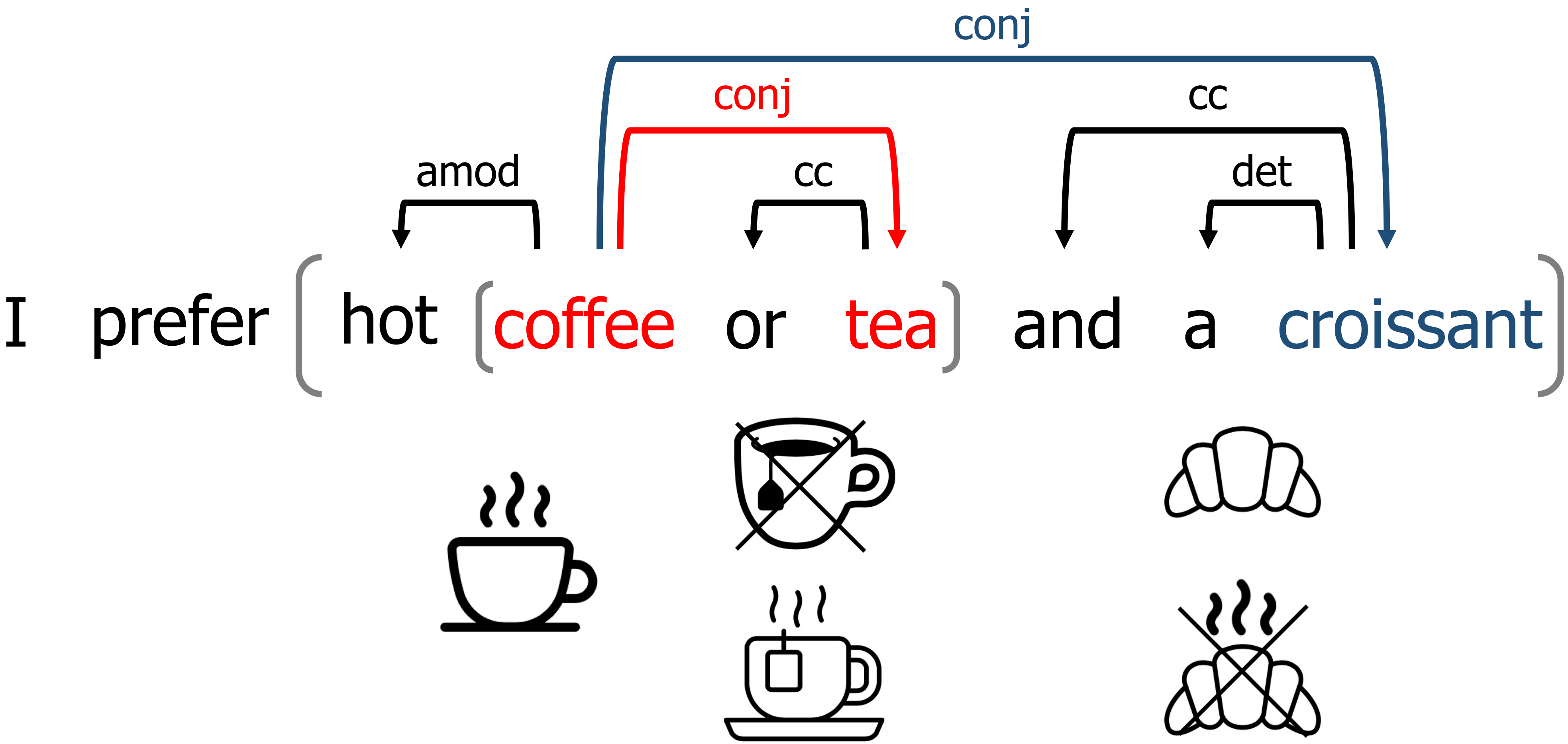


Dependency paradigms exhibit obvious difficulties with coordination because, differently from most linguistic structures, it is not possible to characterize the coordination construct with a general schema involving a head and some modifiers of it. The conjunction itself, has distributional properties 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.

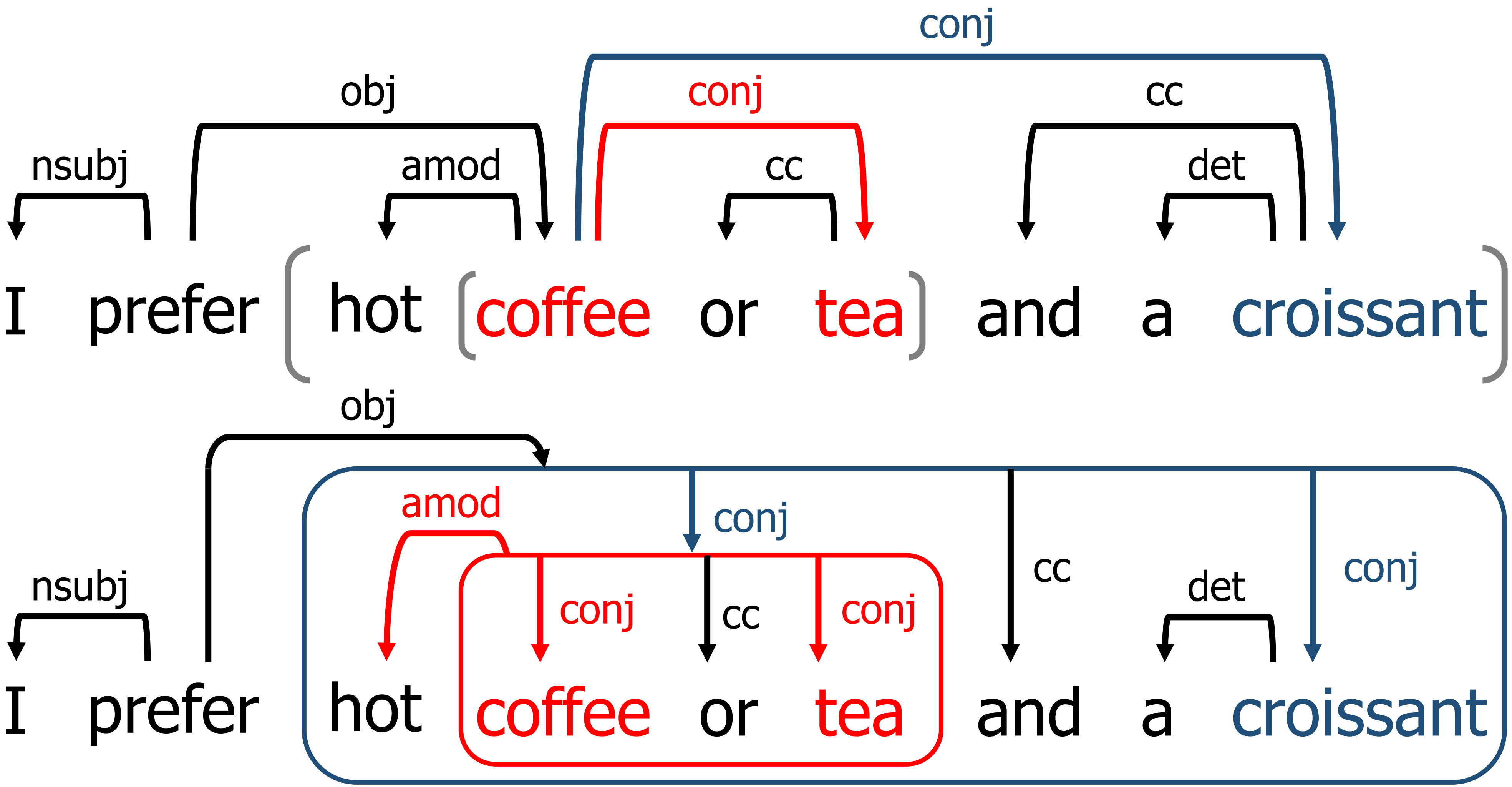
Adding Coordination Boundaries



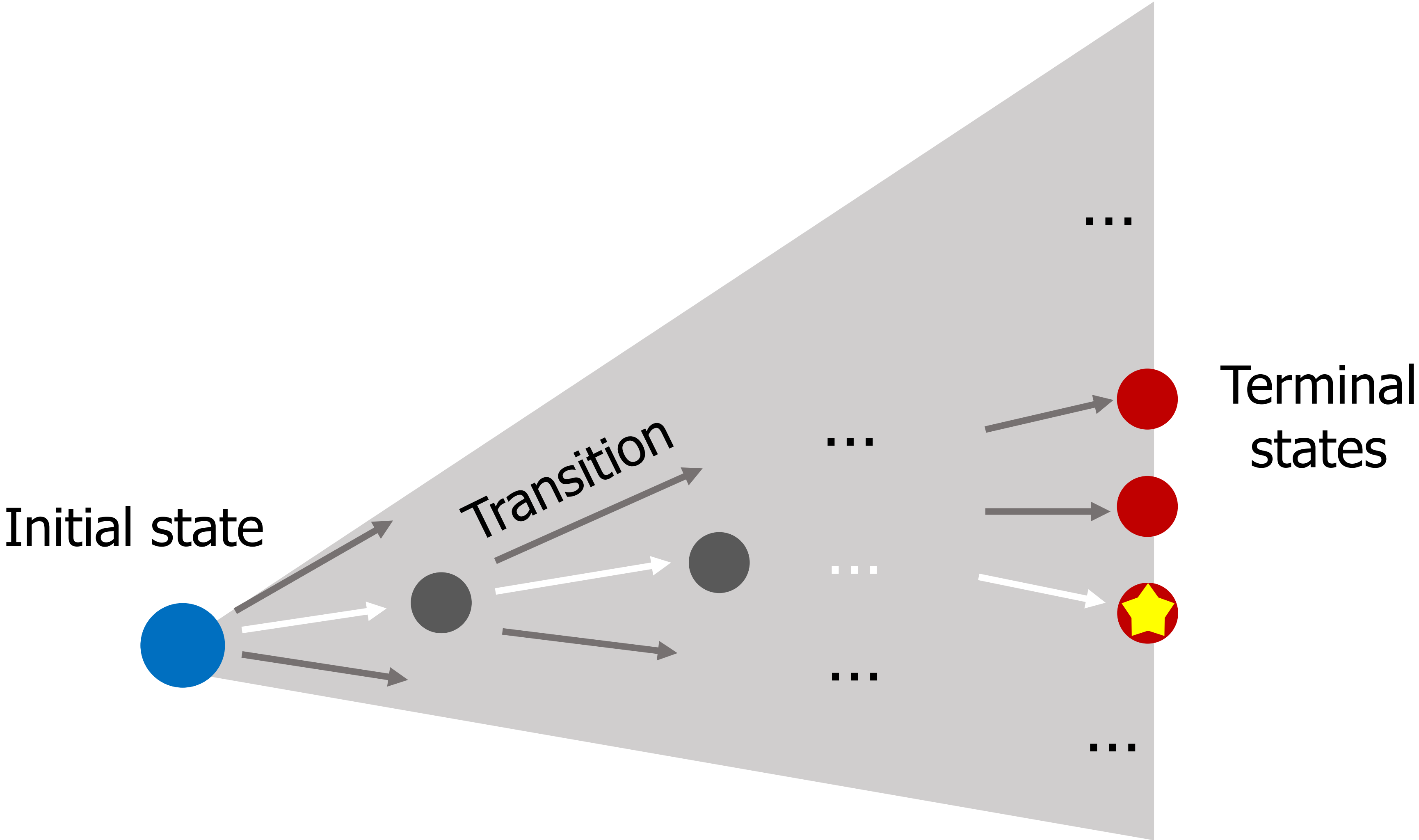
Adding Coordination Boundaries



Bubble Trees (Kahane, 1997)



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

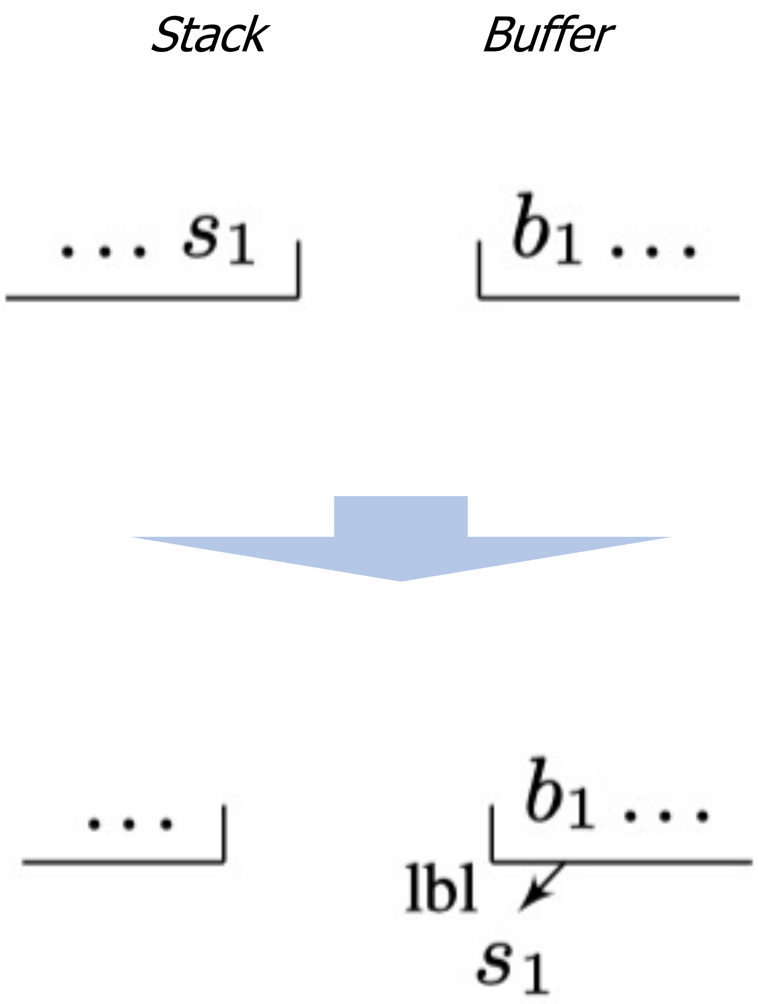
Bubble-Hybrid Transition System

SHIFT

	Stack	*Buffer*
before	$\ldots$	$b_1 \ldots$
after	$\ldots b_1$	$\ldots$

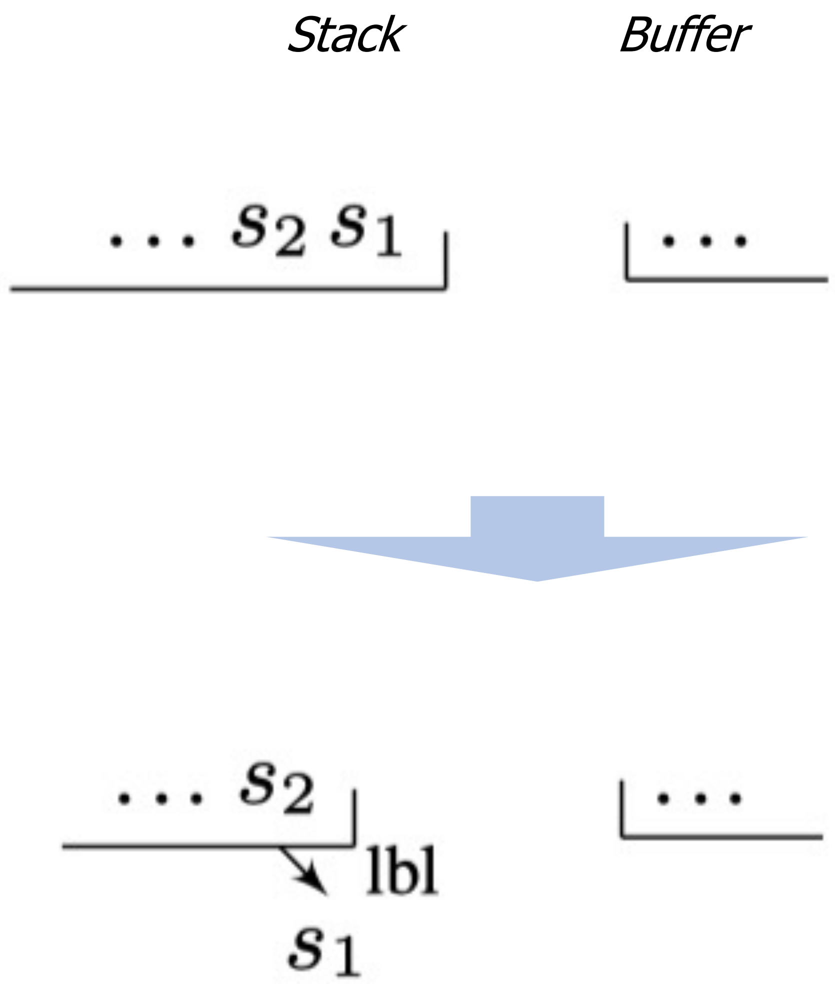
Bubble-Hybrid Transition System

LEFTARC$_{lbl}$



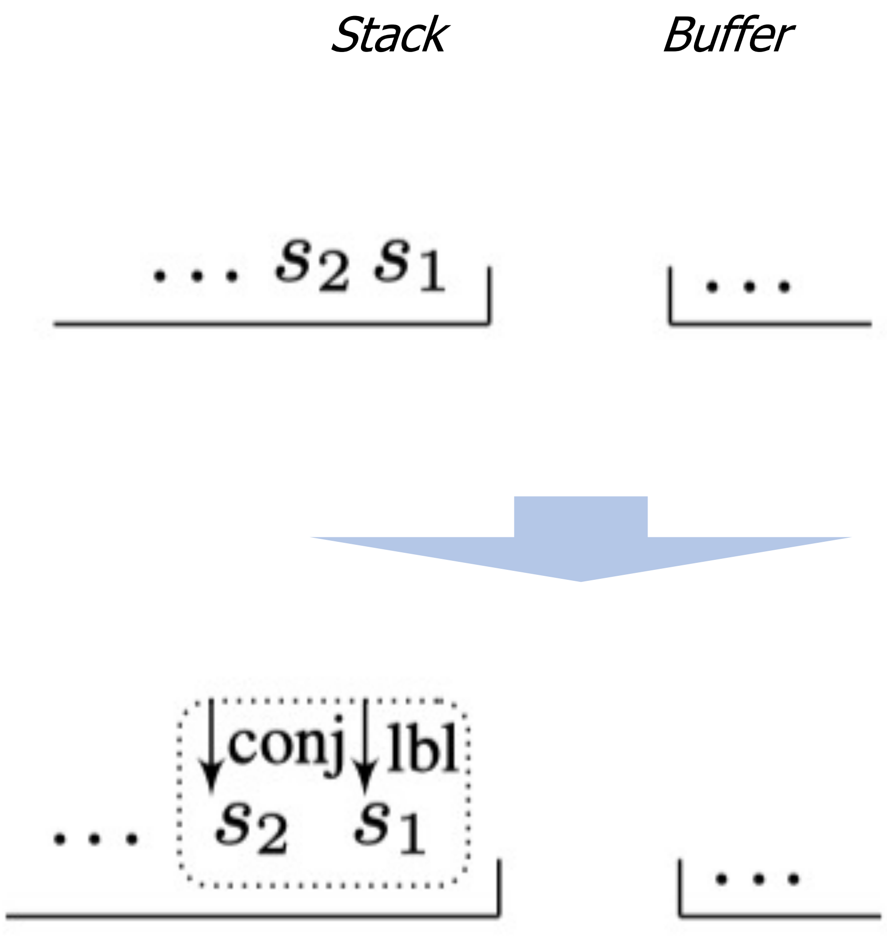
Bubble-Hybrid Transition System

RIGHTARC$_{\text{lbl}}$



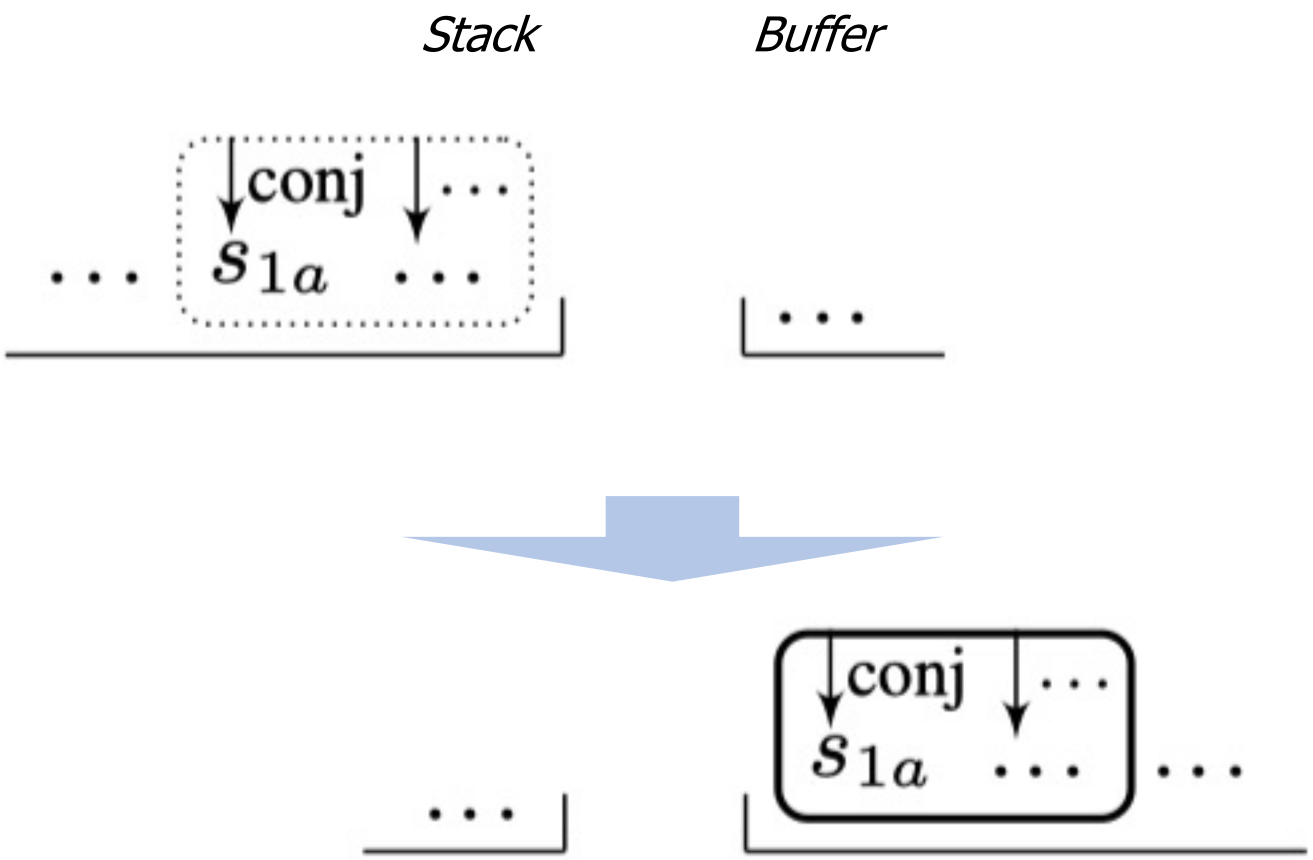
Bubble-Hybrid Transition System

BUBBLEOPEN$_{\mathrm{lbl}}$

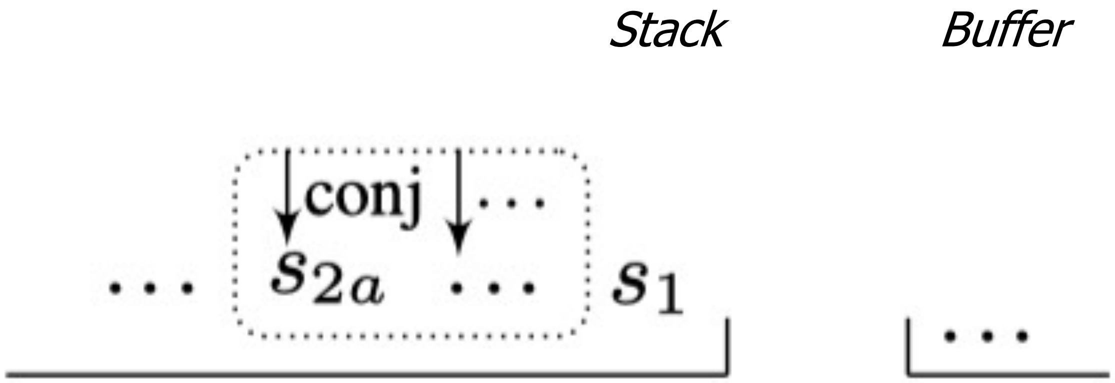


Bubble-Hybrid Transition System

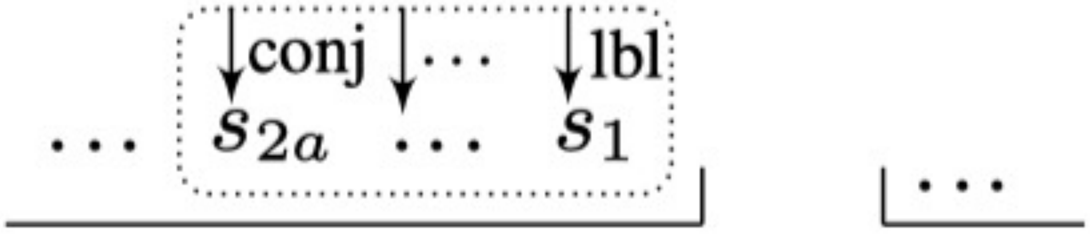
BUBBLECLOSE



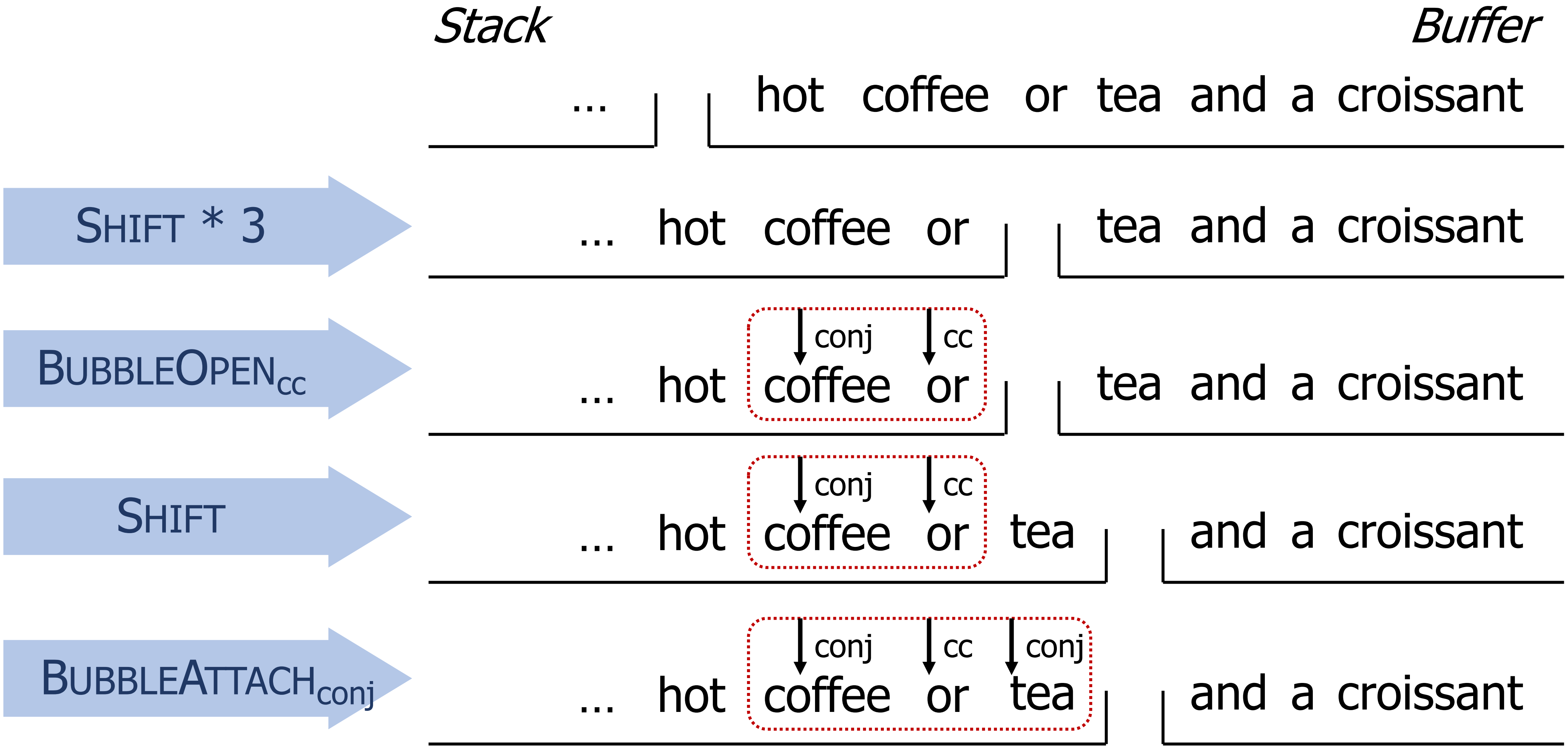
Bubble-Hybrid Transition System



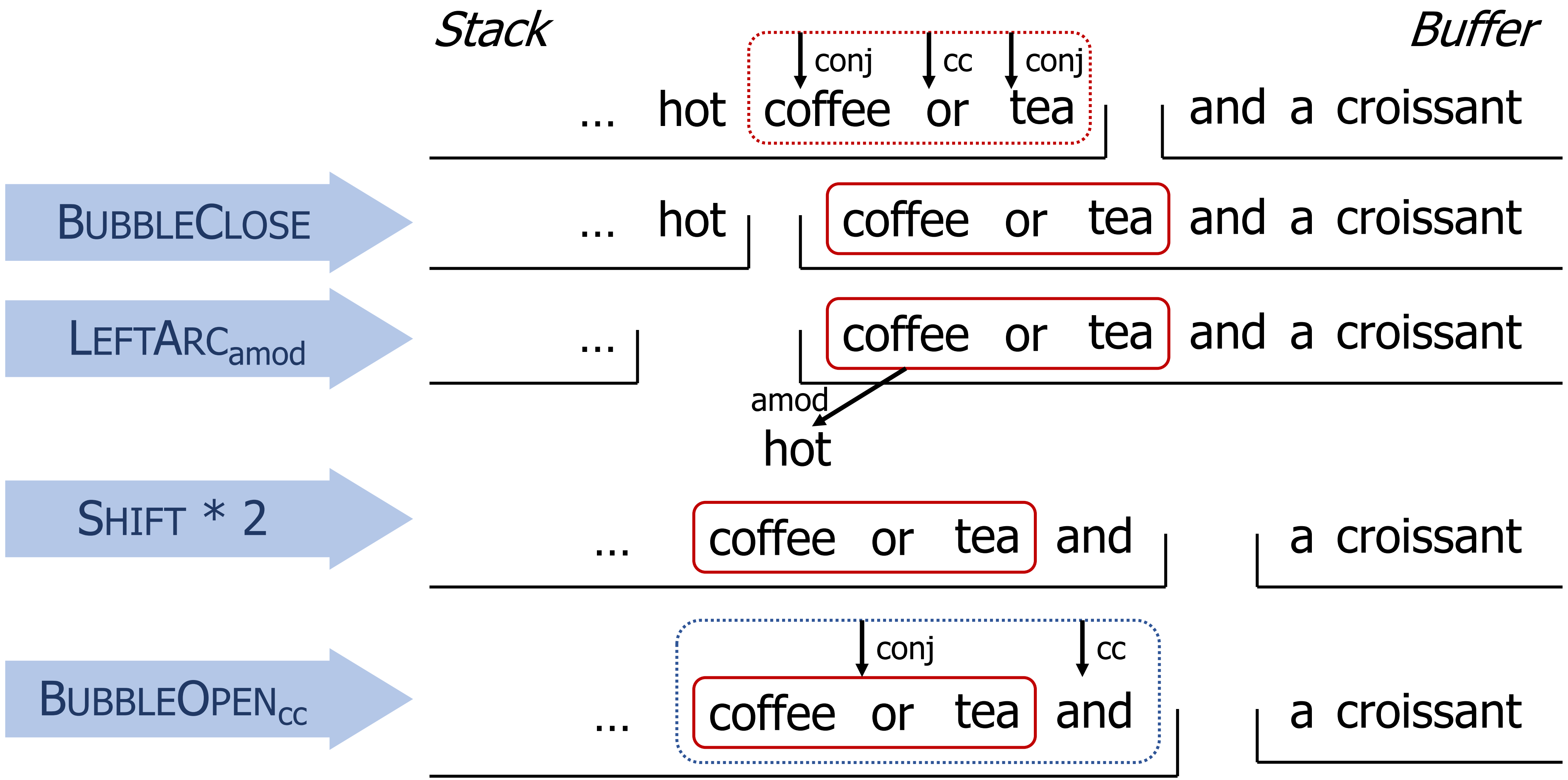
BUBBLEATTACH$_{\text{lbl}}$



Walkthrough of an Example Sentence

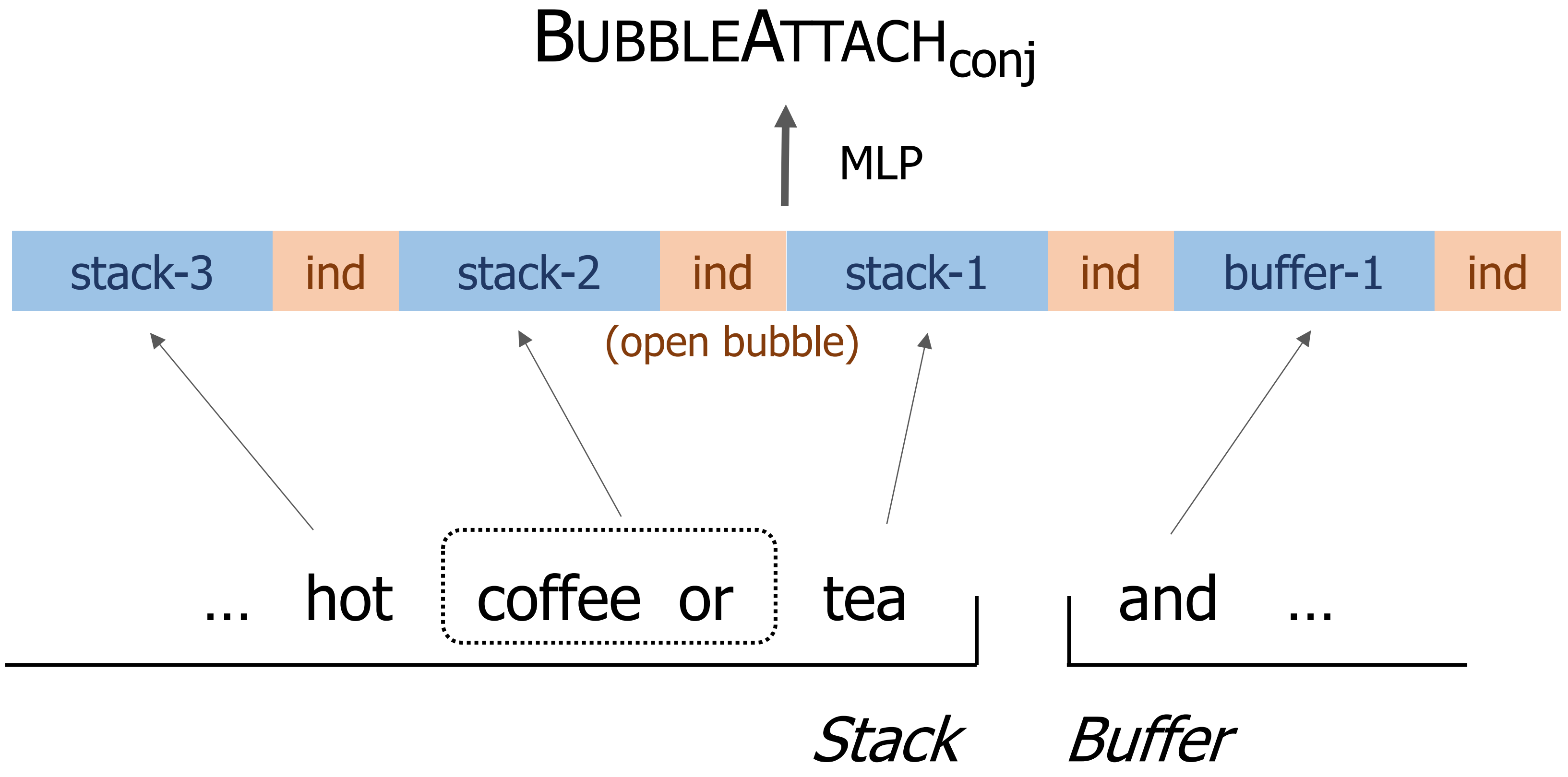


Walkthrough of an Example Sentence

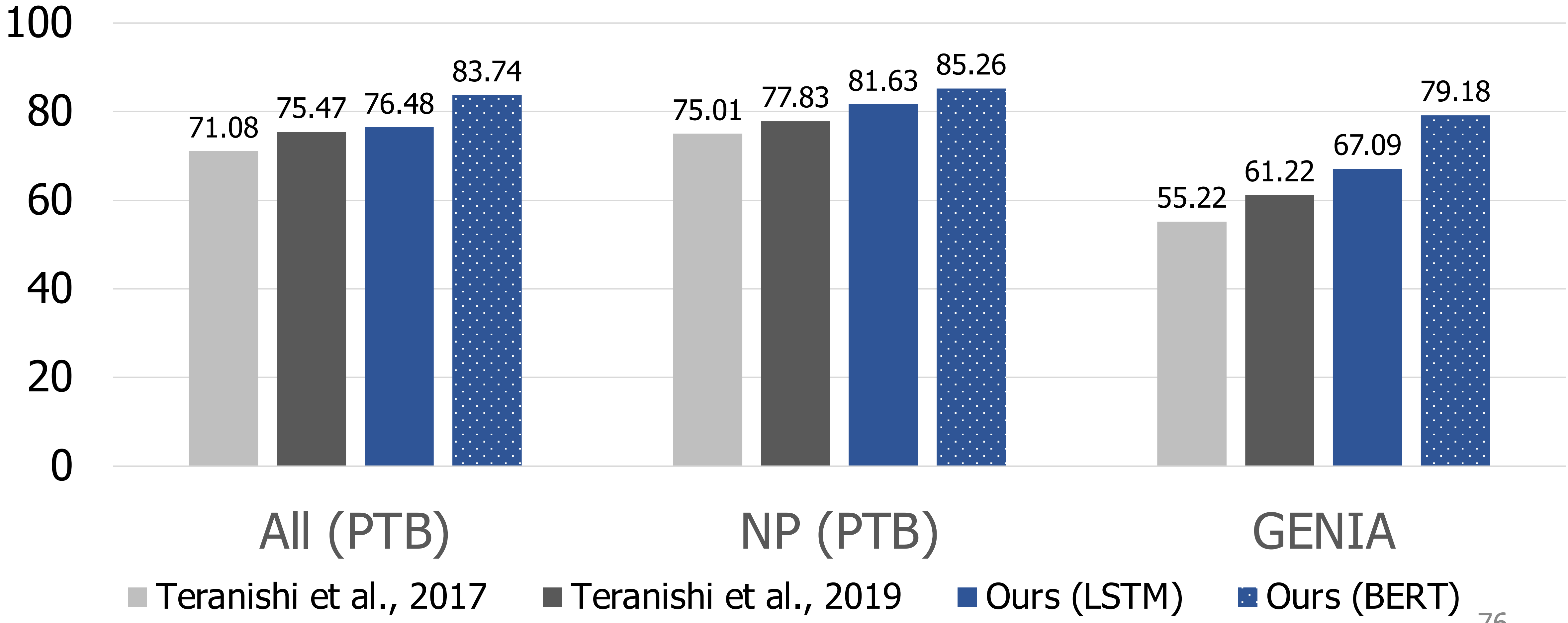


Modeling

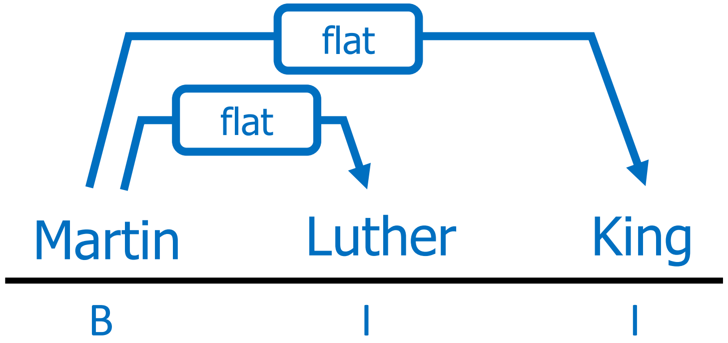
- Follows Kiperwasser and Goldberg's (2016) parser + Greedy Decoder

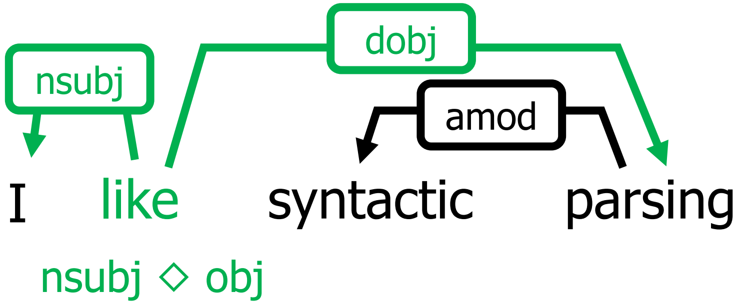


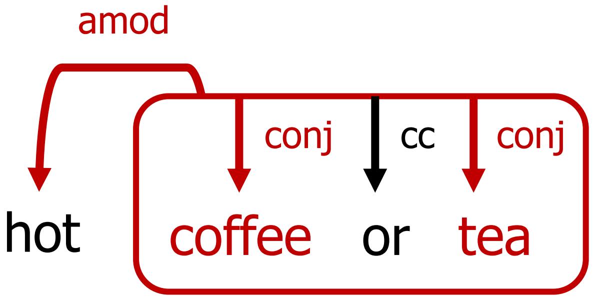
Experiment Results



Summary







Syntactic Phenomenon	Headless MWEs	Core argument structures	Coordination
Involved Relation Types	flat	{nsubj, obj, iobj, csubj, xcomp, ccomp} and more	conj, cc
Constraint / Desired Property	Representational constraint	Valency patterns	Symmetry among conjuncts; Marked coordination boundaries
Proposed Method	Joint tagging and parsing	Joint tagging and parsing	Tree-graph integration; Bubble parsing
Output Structure	Augmented trees	Augmented trees	Beyond trees

Limitations and Future Work

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

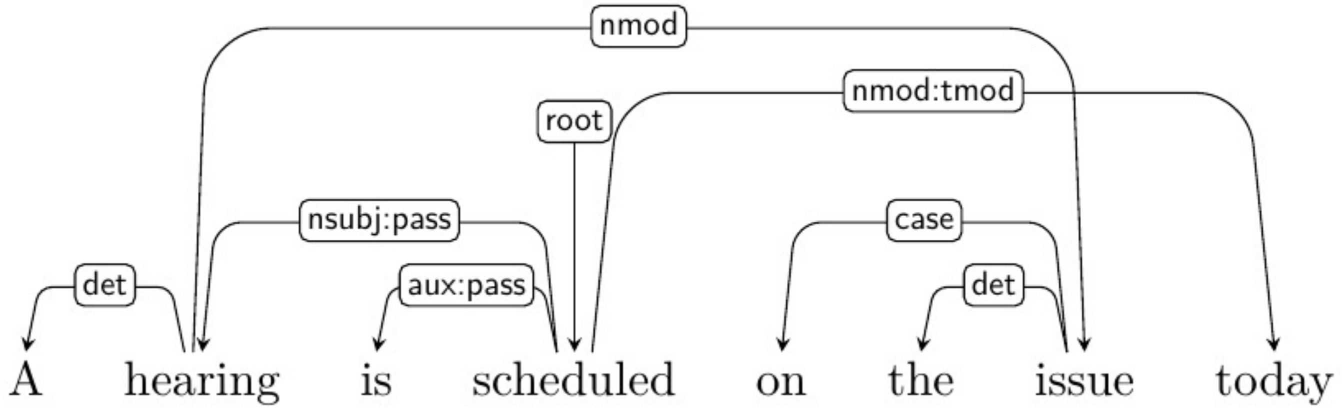
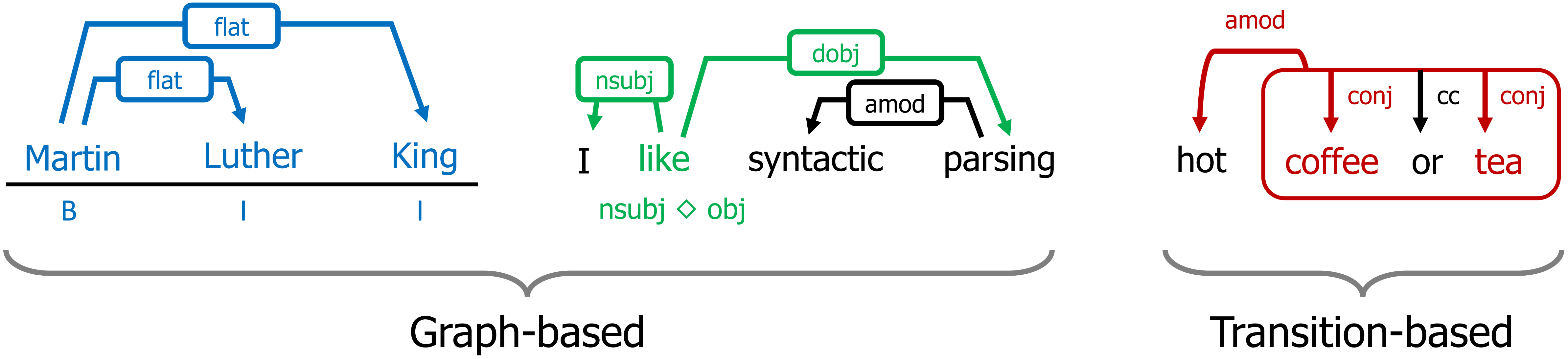


Figure source: Daniel Hershcovich

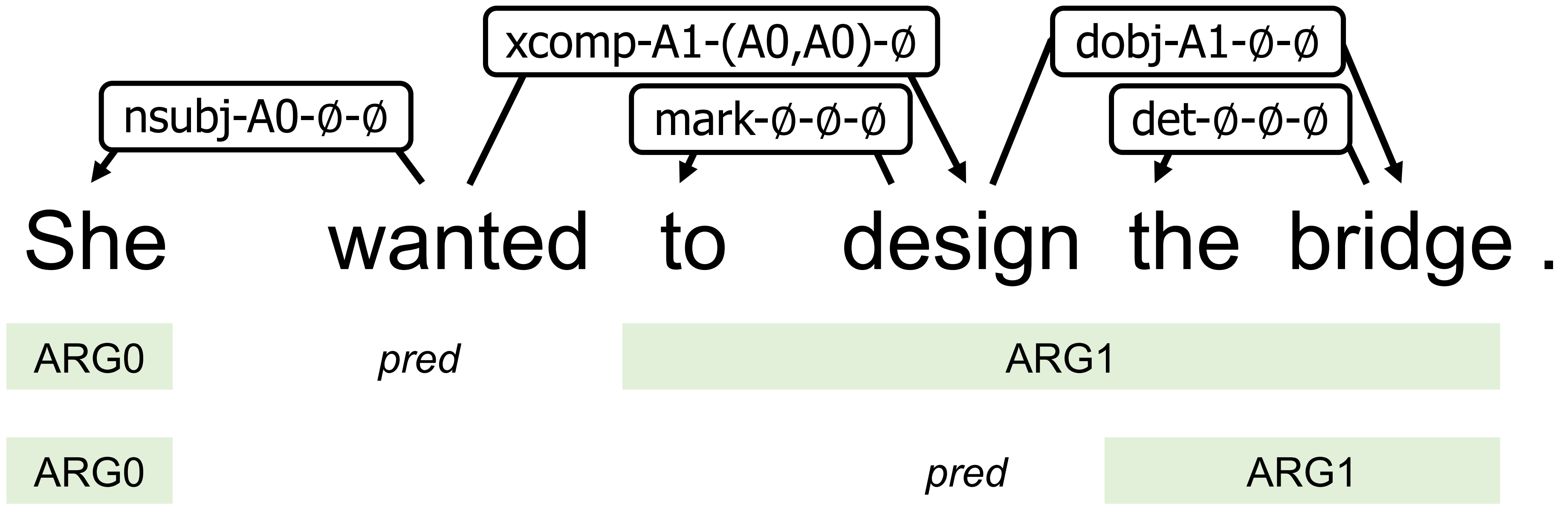
Limitations and Future Work

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



Limitations and Future Work

- Extrinsic evaluation on downstream tasks
  - Previous work: Shi, Malioutov, and İrsoy (EMNLP, 2020)



Universal Dependencies Taxonomy

	Nominals	Clauses	Modifier words	Function words
Core arguments	nsubj, obj, iobj	csubj, ccomp, xcomp		
Non-core dependents	obl, vocative, expl, dislocated	advcl	advmod, discourse	aux, cop, mark
Nominal dependents	nmod, appos, nummod	acl	amod	det, clf, case
Coordination	**MWE**	**Loose**	**Special**	**Others**
conj, cc	fixed, flat, compound	list, parataxis	orphan, goeswith, reparandum	punct, root, dep

All About Parsing