自然语言处理

2022年秋季

黄河燕、鉴萍
北京理工大学 计算机学院
hhy63, pjian@bit.edu.cn

NLP中的序列结构预测

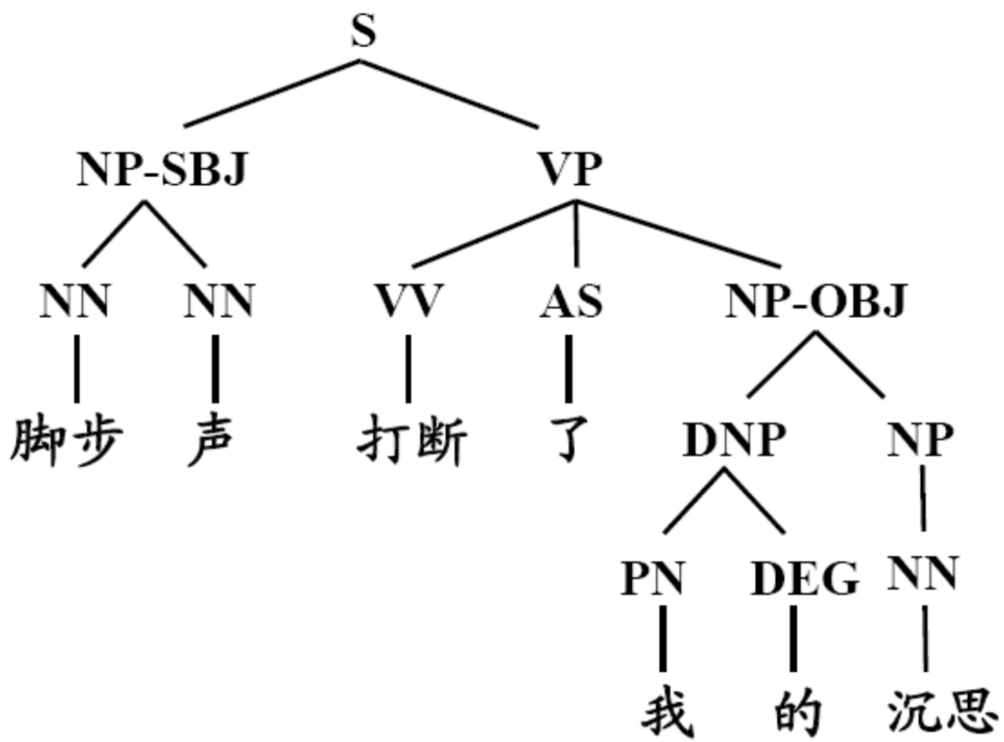
黄河燕，鉴萍
北京理工大学 计算机学院
hhy63, pjian@bit.edu.cn

大纲

- 句法分析
  - 短语结构句法分析
  - 依存结构句法分析
  - 浅层句法分析
- 篇章分析
  - 篇章结构分析

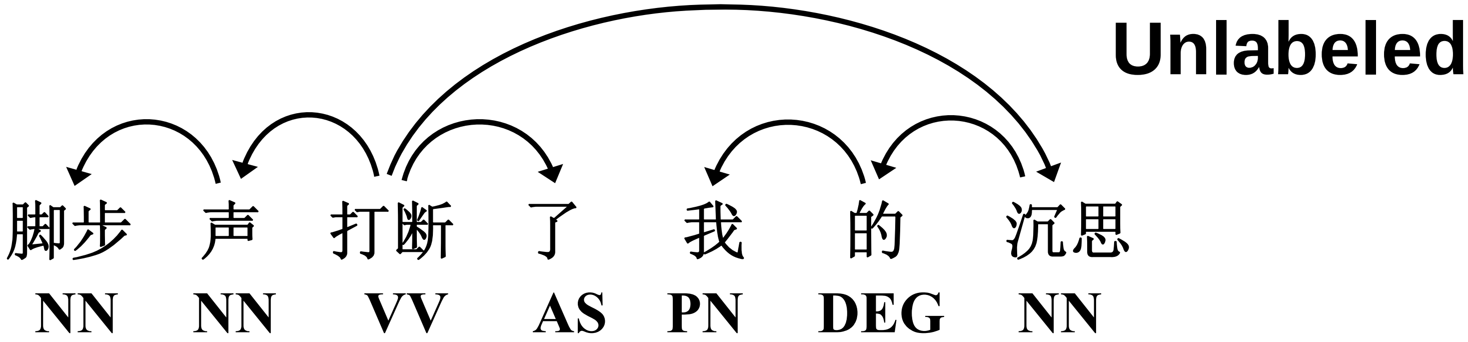
依存结构句法分析

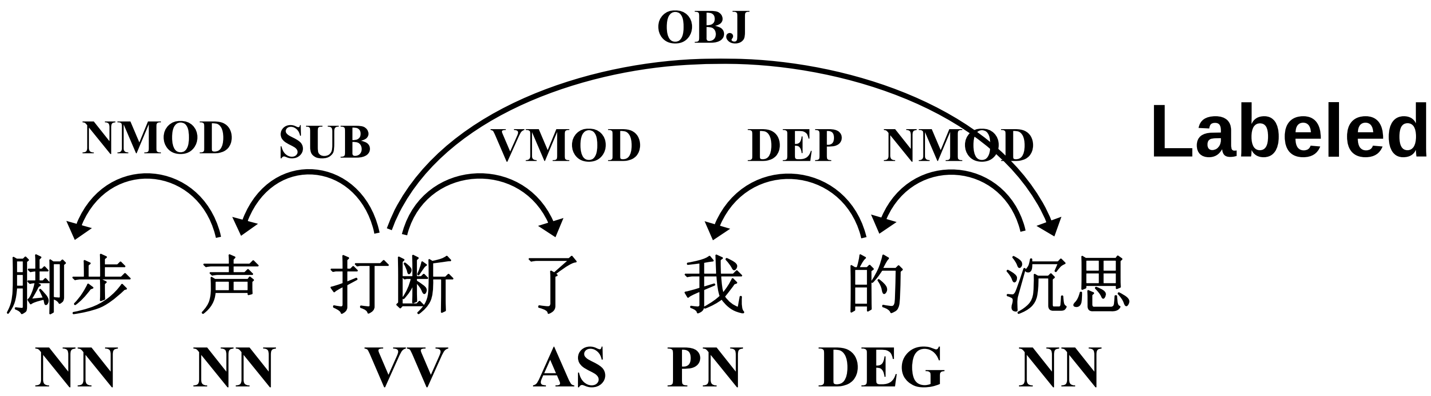
◘ 短语结构



依存结构句法分析

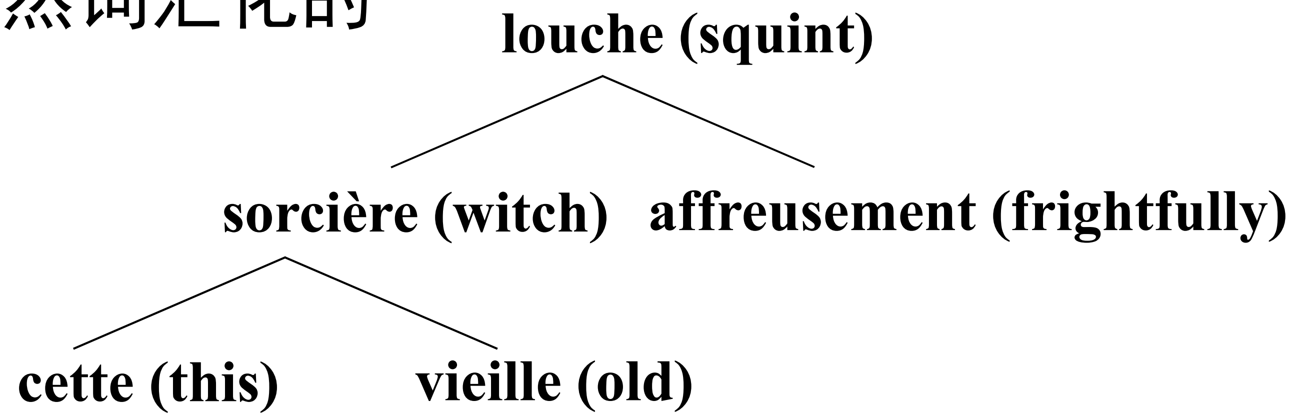
依存结构





依存结构句法分析

- 依存文法，dependency grammar (grammaire de dépendance, Lucien Tesnière, 1893-1954)
  - 用词之间的支配和被支配关系来描述语言结构
  - 天然词汇化的

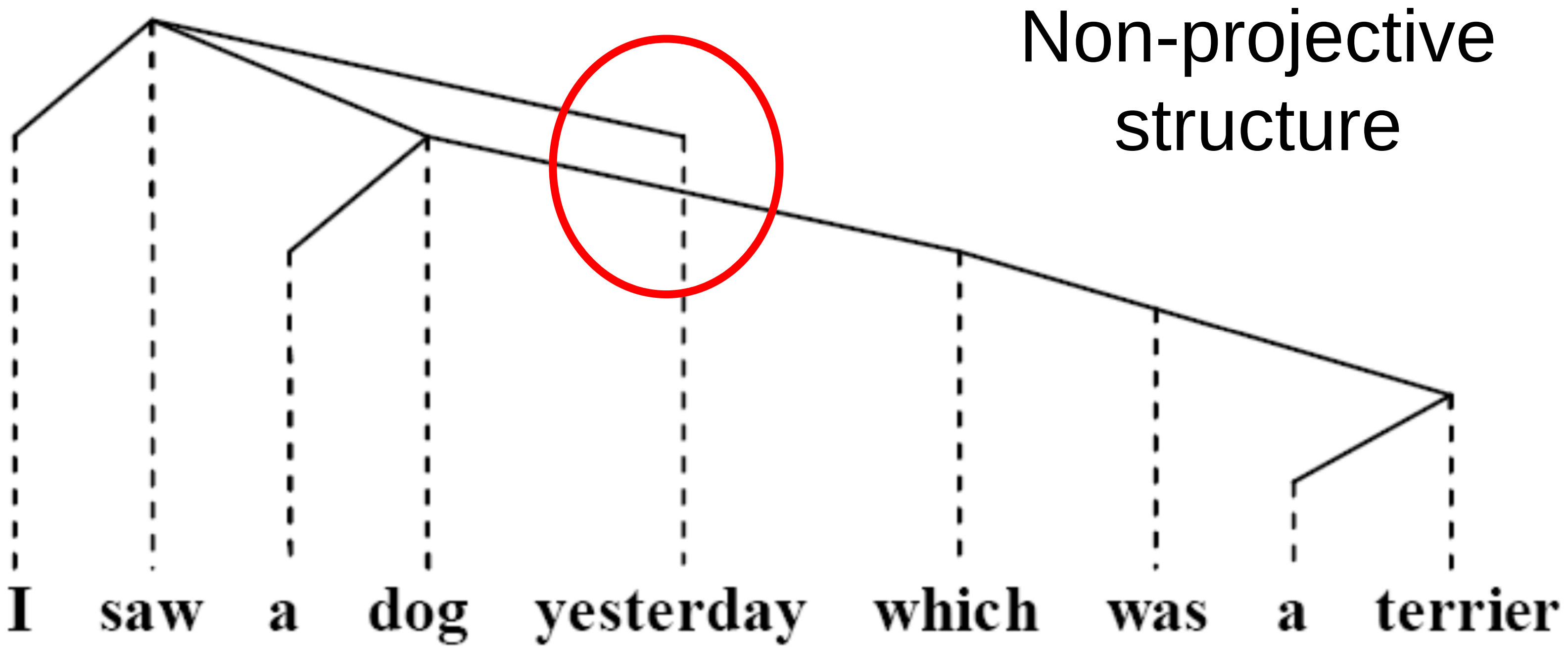


Cette sorcière vieille louche affreusement

这个老巫婆可怕地眯着眼

依存结构句法分析

- 两个优点
  1. 相比短语结构，更适合free order语言

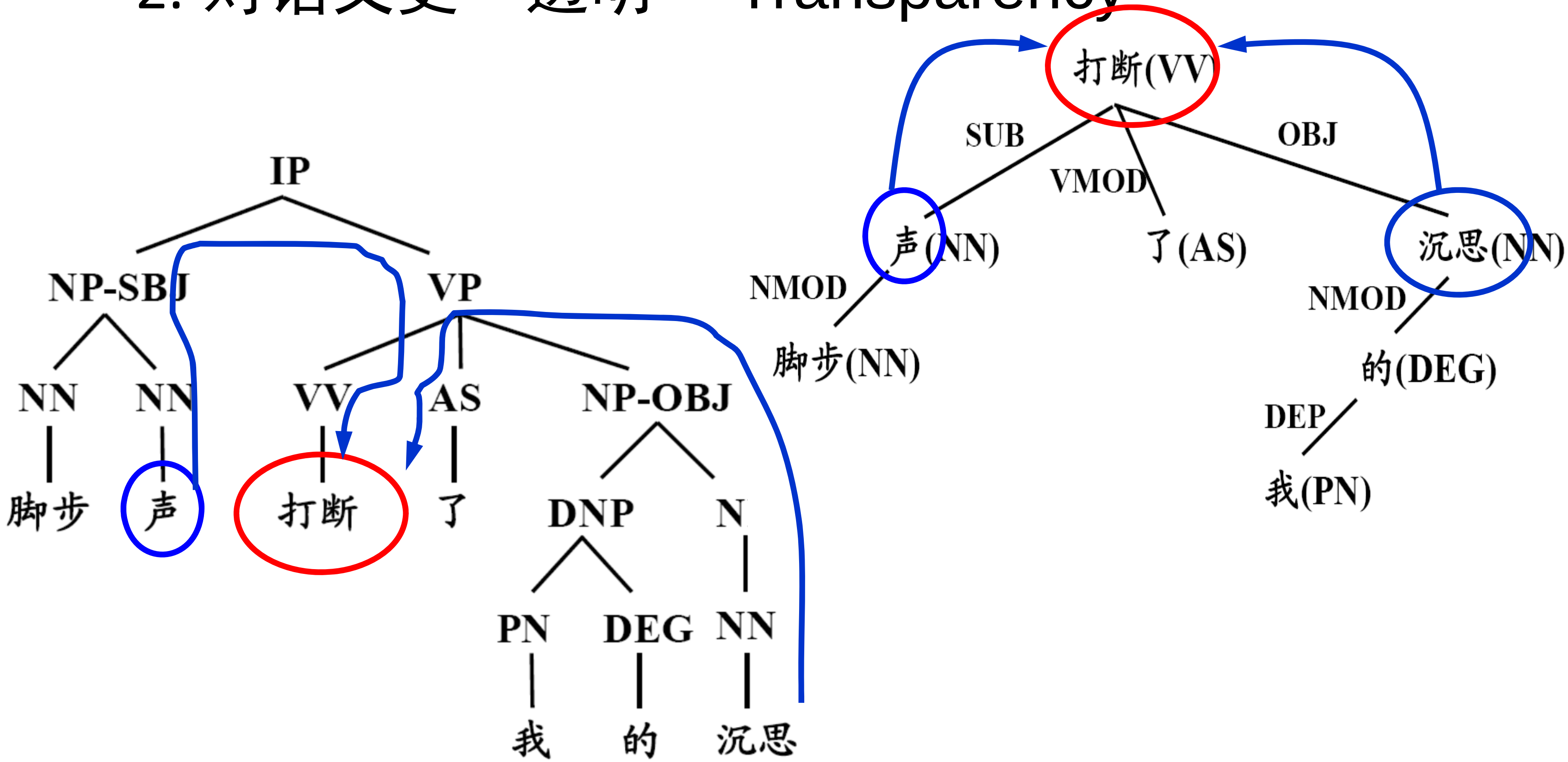


free order语言会出现**Non-projective**结构，相比短语结构分析，依存分析更易处理这种特殊结构

依存结构句法分析

两个优点

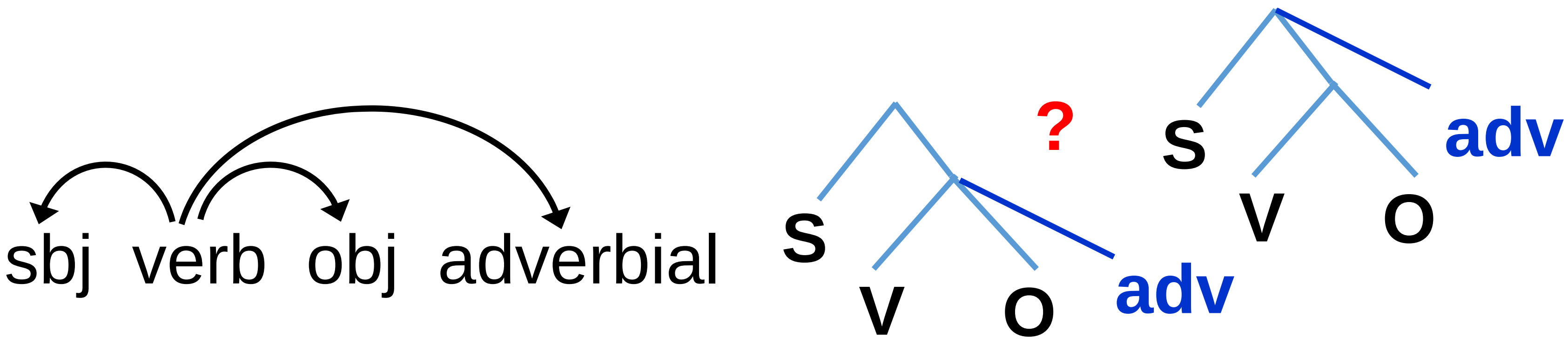
2. 对语义更“透明” Transparency



依存结构句法分析

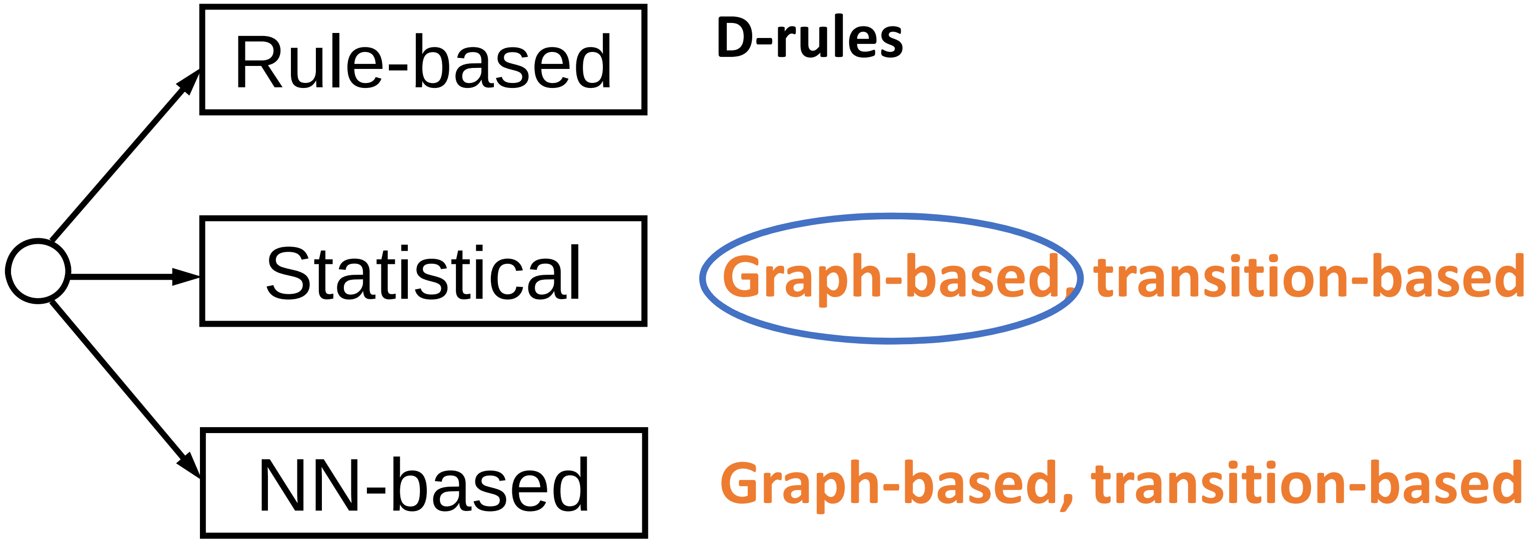
与短语结构之间的转化

- 短语结构+headrules ⇒ 依存结构
- 依存结构+组合的优先顺序 ⇒ 短语结构



依存结构句法分析

◘ 依存分析方法



因为缺少形式化(**grammar**)，依存分析主要在统计框架下发挥了它的优势

依存结构句法分析

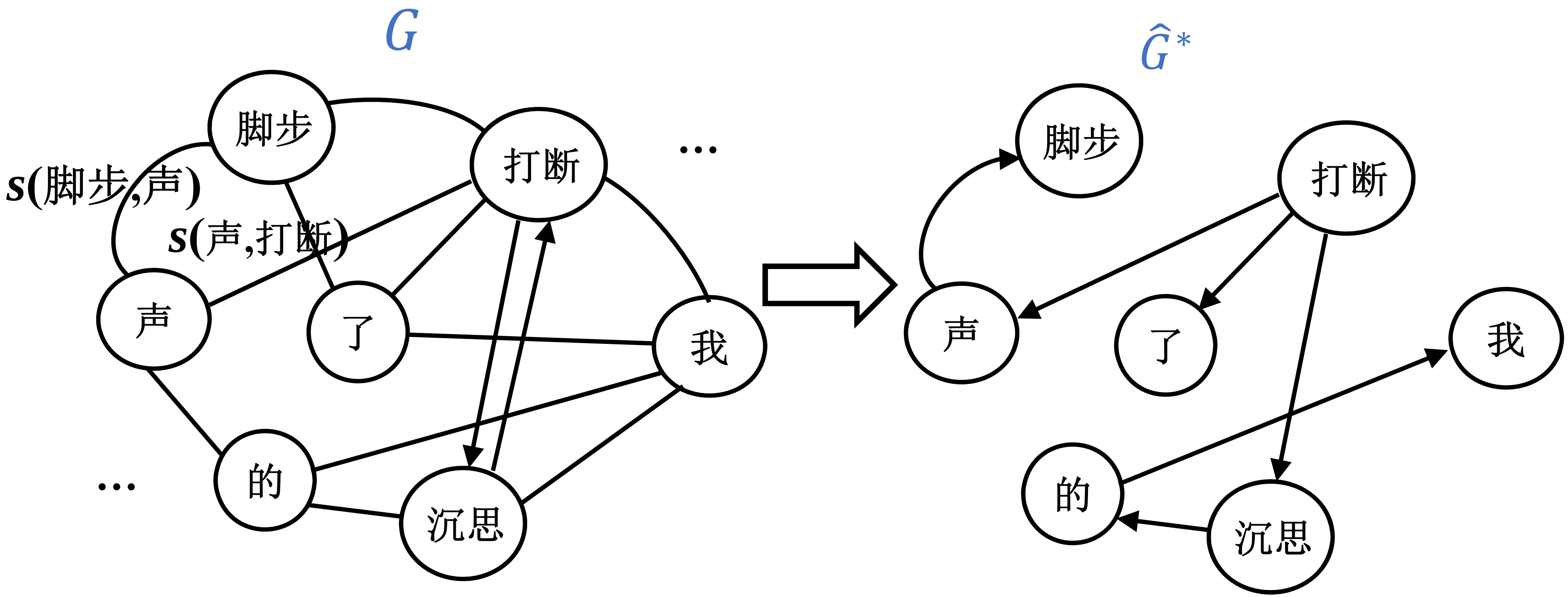
◻ Graph-based (基于图的模型)

- MST model (maximum spanning tree model, 最大生成树模型)

——将依存树结构看作一个生成树，求打分最大的那个 $$\hat{G}^* = \underset{\hat{G} \in T(G)}{\operatorname{argmax}} \operatorname{Score}(\hat{G})$$

$T(G)$：图G的所有生成树

Graph-based



生成树：包含连通图中所有的顶点；任意两顶点之间有且仅有一条通路

两个结点间的**score**代表这两个结点(词)构成依存关系的打分

Graph-based

- 假设一棵生成树的打分/概率/权重由它的边来因子化(factored)：

$$s(\boldsymbol{x}, y) = \sum_{(i,j)\in y} s(i,j) = \sum_{(i,j)\in y} \boldsymbol{w}\cdot\boldsymbol{f}(i,j)$$

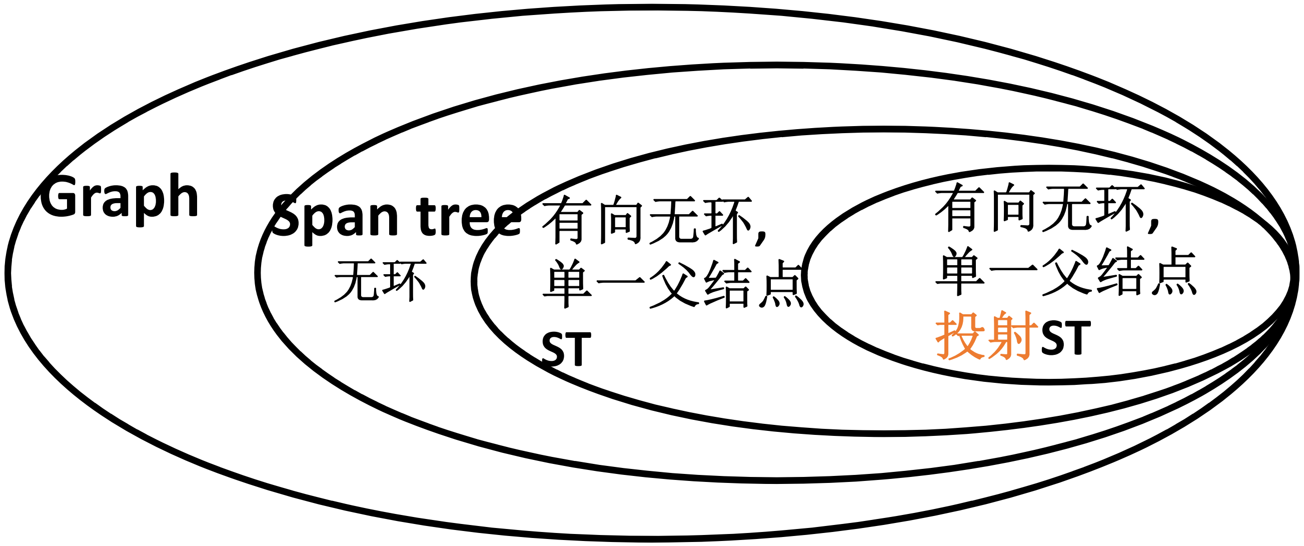
$\boldsymbol{f}$为高维二元特征函数

典型的判别式模型

$$f(i,j) = \begin{cases} 1 & \text{if } x_i =' \text{打断}' \text{and } x_j =' \text{沉思}' \\ 0 & \text{otherwise} \end{cases}$$

Graph-based

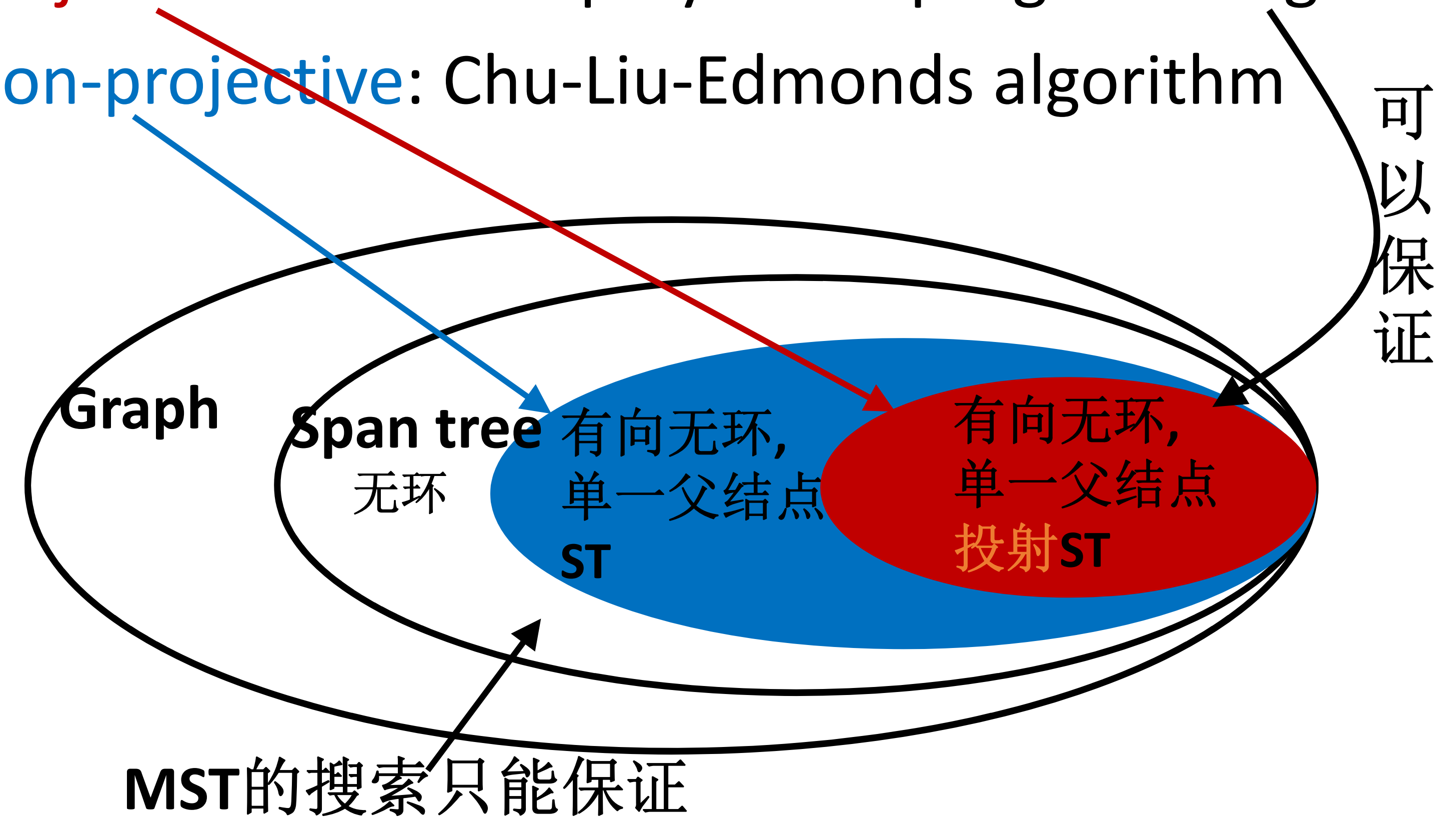
- 如何搜索最大打分的树？(inference)
  - 是一棵“生成树”，不足以保证是某一语言一棵合格的依存树



Graph-based

MST parser只借助了MST的形式，针对不同的依存结构，设计了特定的搜索算法

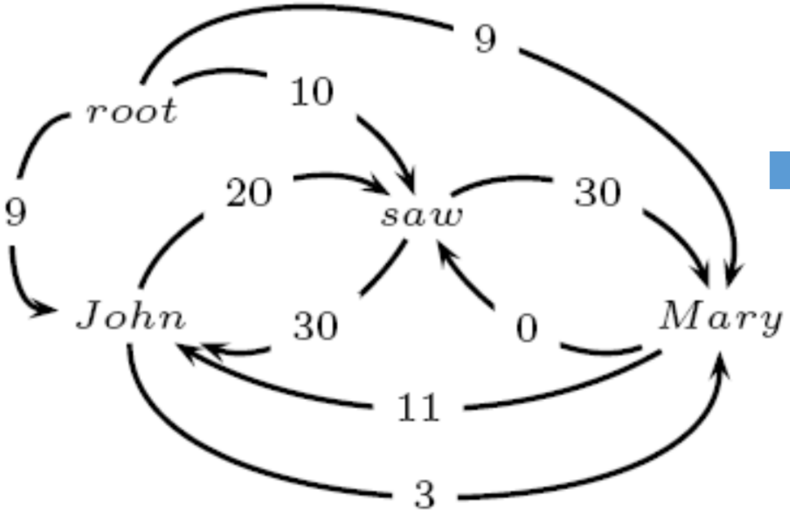
- 如何搜索最大打分的树？(inference)
  - Projective: bottom-up dynamic programming
  - Non-projective: Chu-Liu-Edmonds algorithm



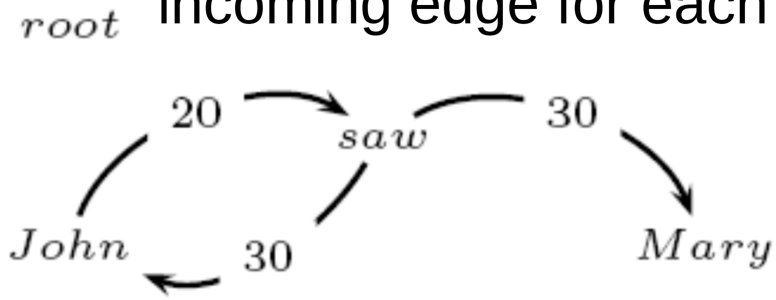
Graph-based

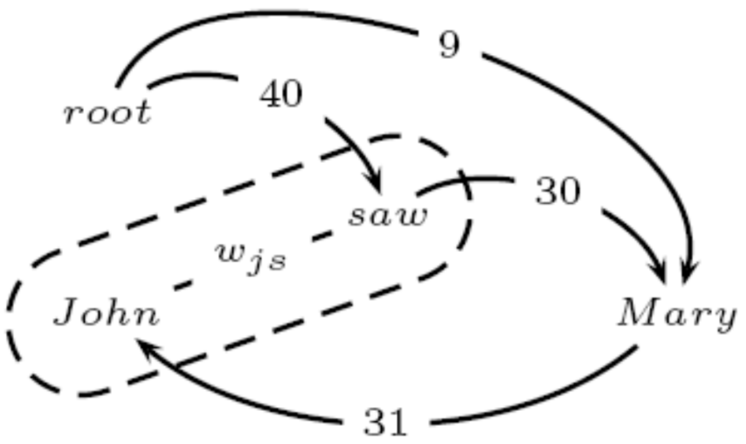
- 如何搜索最大打分的树？(inference)
  - 针对汉语、英语等projective结构来说，自下而上的动态规划算法即可(约束到投射树)
  - 针对non-projective的语言如拉丁语、朝鲜语以及斯拉夫语族如俄语、捷克语等，可采用图上的约束算法如Chu-Liu-Edmonds算法

- Chu-Liu-Edmonds algorithm
  - **Non-projective DT只要保证得到的MST无环而且**单一父结点即可

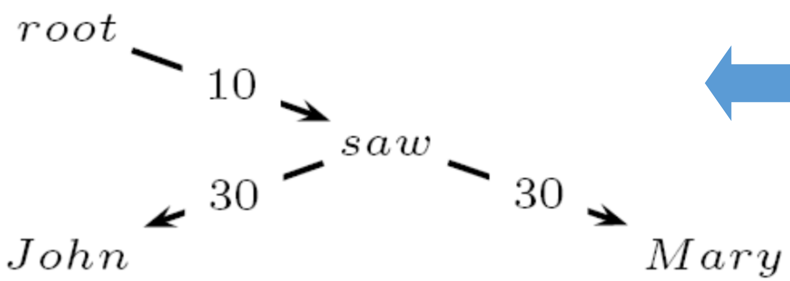


1. Find the highest scoring incoming edge for each word



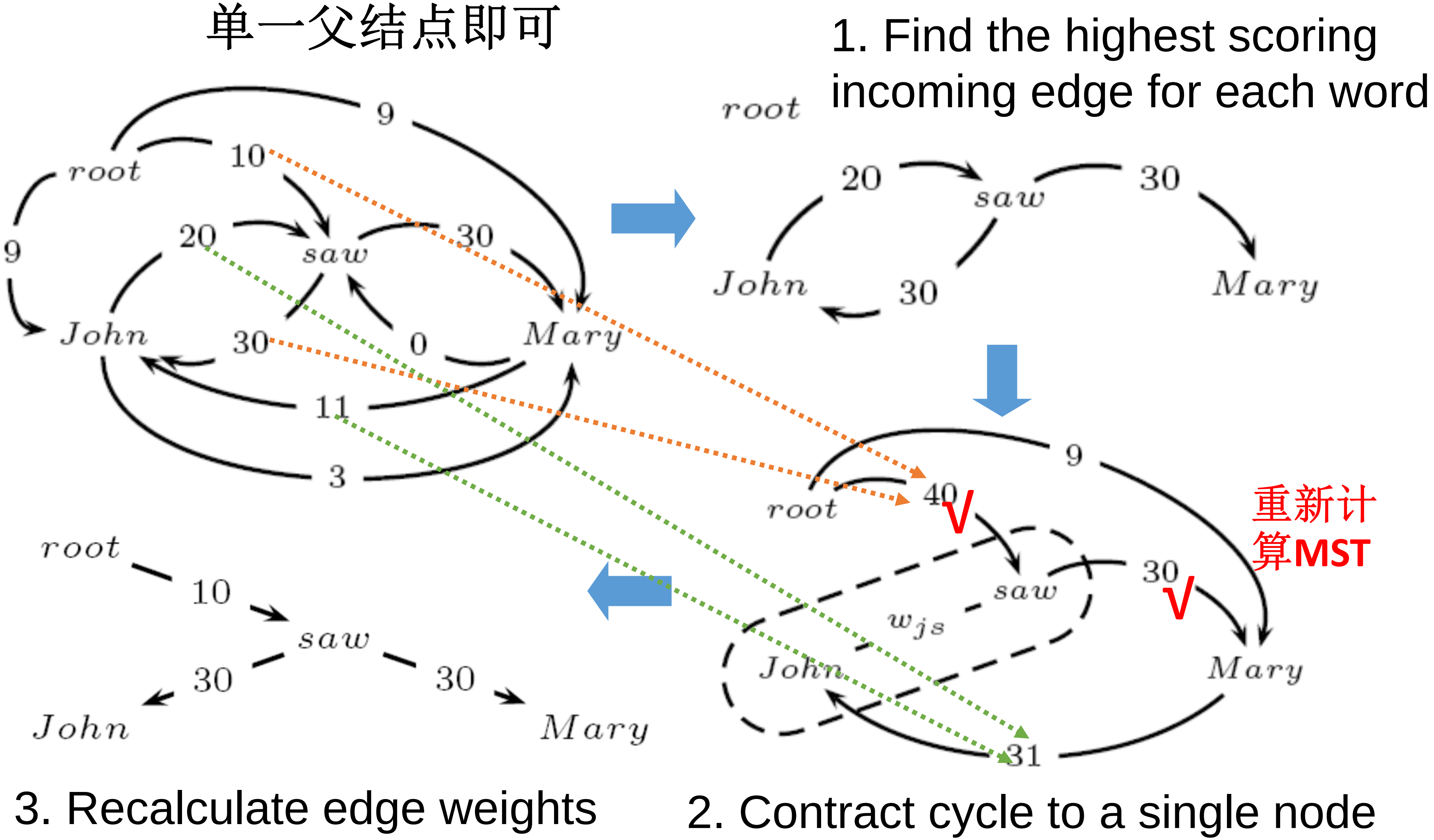


2. Contract cycle to a single node



3. Recalculate edge weights

- Chu-Liu-Edmonds algorithm
  - **Non-projective DT**只要保证得到的**MST**无环而且单一父结点即可



Graph-based

◘ 参数学习采用在线学习方法

- MIRA (margin infused relaxed algorithm)

$$\min \left\| w^{(i+1)} - w^{(i)} \right\|$$
$$\text{s.t. } s(x_t, y_t) - s(x_t, y') \geq L(y_t, y')$$
$$\forall y' \in best_k\left(x_t; w^{(i)}\right)$$

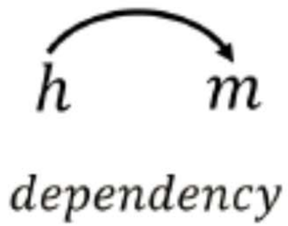
考虑到**margin**至少与错误分类的损失一样大时，
保持权重向量更新尽可能小

——以最小的参数更新获得较大的**score**提升

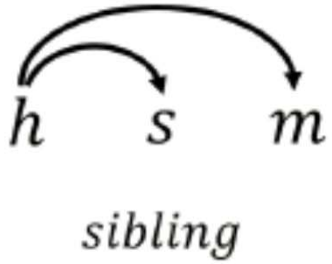
Graph-based

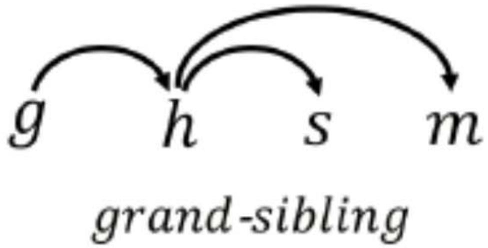
- 高阶MST

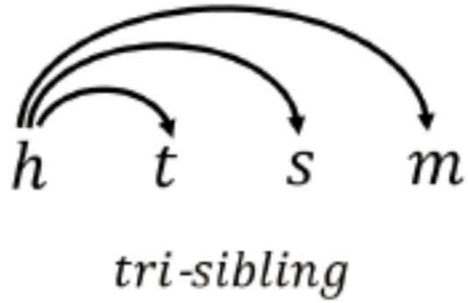
First order factorization



Second order factorization







Third order factorization

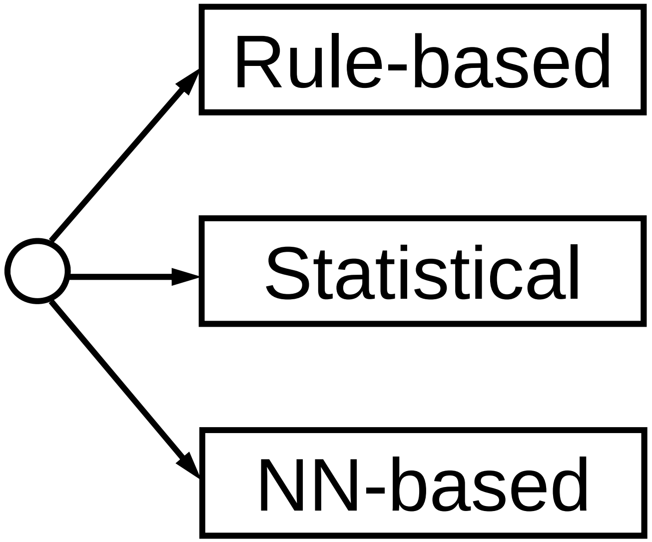
Graph-based

□ Graph-based方法优缺点

- 判别式模型，直接为条件概率建模(准确率高)
- 全局寻优，时间复杂度$O(n^3)$或$O(n^2)$
- 将projective和non-projective统一在一个框架
- 不易使用动态特征
- 不能直接预测relation type

依存结构句法分析

◘ 依存分析方法



D-rules

Graph-based, transition-based

Graph-based, transition-based

依存结构句法分析

◻ Transition-based (基于转移的方法)

一种确定性的分析方法，将句法结构转化为一系列的动作，动作导致的状态转移最终规约出句法树

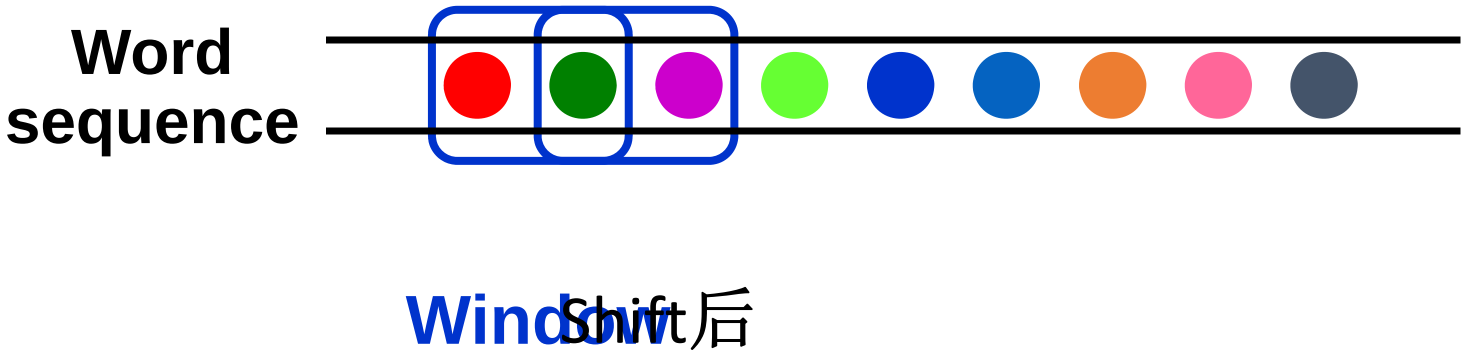
$$t^* = \operatorname{argmax} \sum_{t \in T} \operatorname{score}(c, t)$$

t: transition
c: configuration(可以看作当前状态)

- Yamada's parser
- Nivre's parser

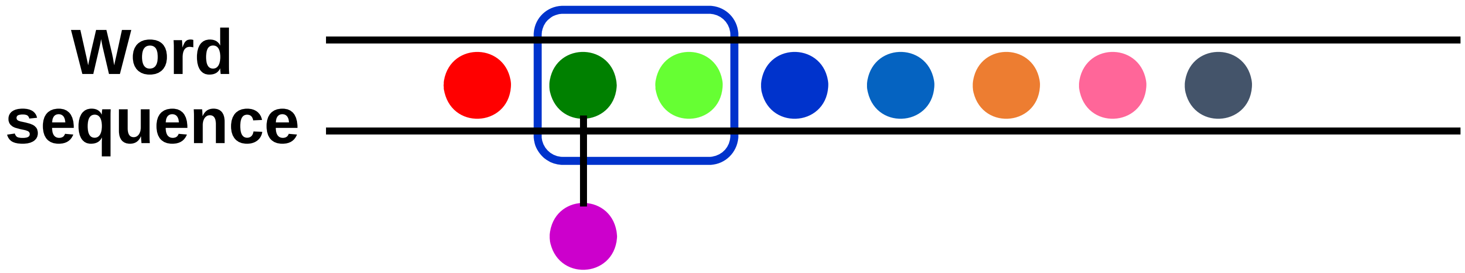
Transition-based

- Yamada’ s parser (Yamada, 2003)
  - Multi-pass shift-reduce parser
  - 动作集合: Shift, Right, Left



Transition-based

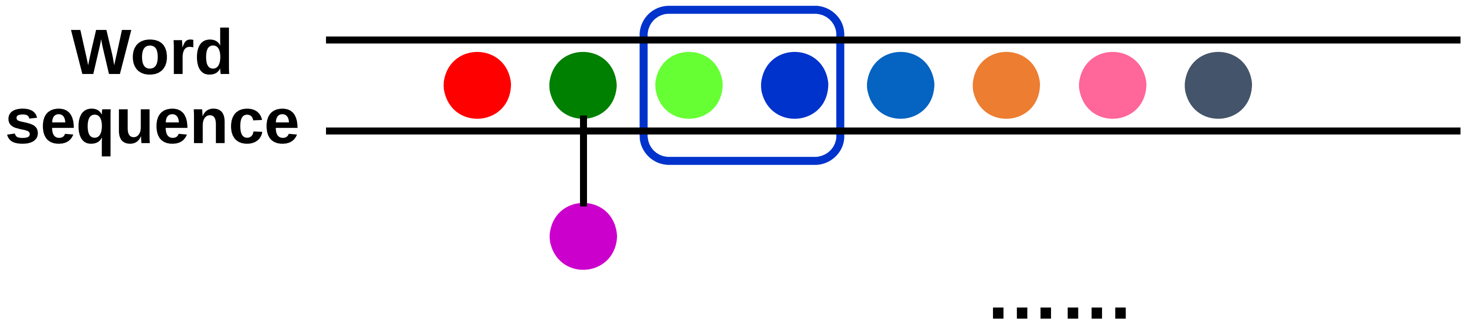
- Yamada' s parser (Yamada, 2003)
  - Multi-pass shift-reduce parser
  - 动作集合: Shift, Right, Left



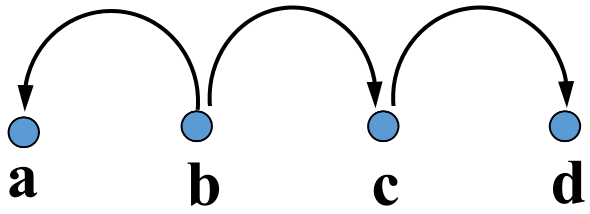
Left后

Transition-based

- Yamada' s parser (Yamada, 2003)
  - Multi-pass shift-reduce parser
  - 动作集合: Shift, Right, Left



Shift后



Pass1: [a b]→Right→Shift→Shift→Left→

Pass2: [a b]→Left

初始化: $[\ ]_W$

结束: $[\mathrm{ROOT}]_W$

关注window中的两个token

Right

$$\frac{\ldots [w_i, w_{i+1}]_W \ldots}{\ldots [w_{i+1}]_W \ldots \qquad A \cup \{w_i \leftarrow w_{i+1}\}}$$

w_i

W: 窗口

i: 词的位置

A: 弧的集合

Left

$$\frac{\ldots [w_i, w_{i+1}]_W \ldots}{\ldots [w_i, w_{i+2}]_W \ldots \qquad A \cup \{w_i \rightarrow w_{i+1}\}}$$

w_{i+1}

$$\frac{\text{动作执行之前}}{\text{动作执行之后}}$$

Shift

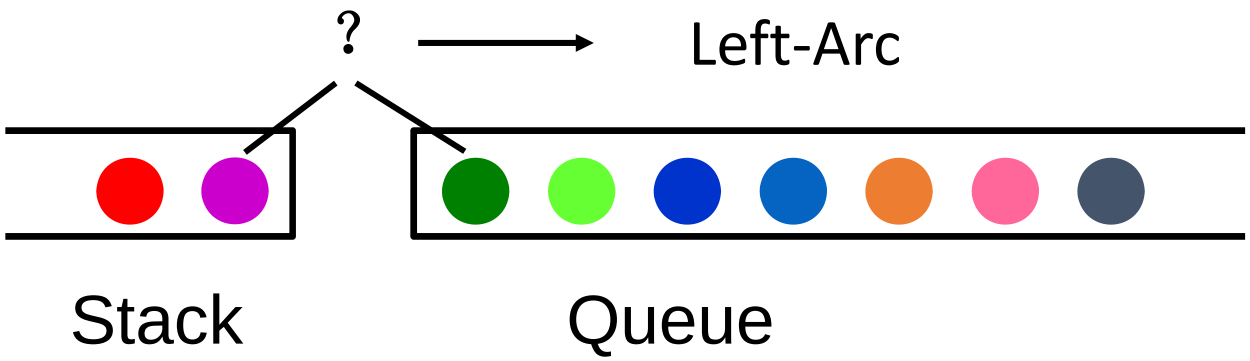
$$\frac{\ldots [w_i, w_{i+1}]_W \ldots}{\ldots [w_{i+1}, w_{i+2}]_W \ldots}$$

Actions in Yamada's parser

Transition-based

Nivre' s parser (Nivre, 2003)

- One-pass shift-reduce parser
- 两个模型:
  - **Arc-standard:** Left-Reduce, Right-Reduce, Shift
  - **Arc-eager:** Left-Arc, Right-Arc, Reduce, Shift(只介绍)

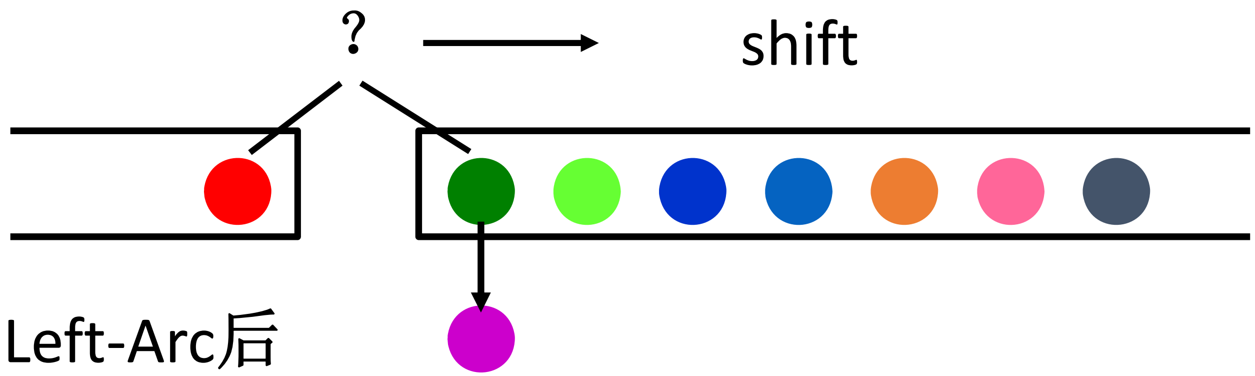


Arc-eager每次关注栈顶的**token**和输入队列的第一个**token**

Transition-based

Nivre’ s parser (Nivre, 2003)

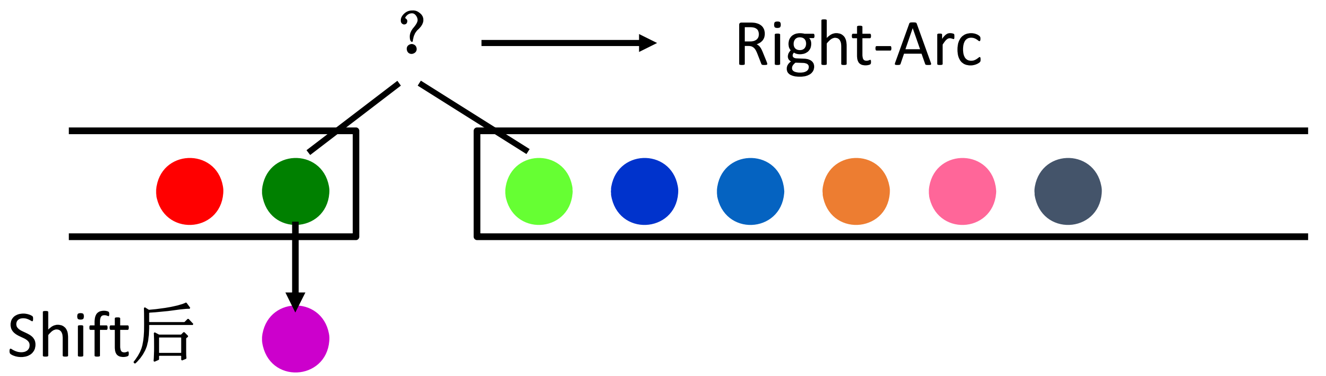
- One-pass shift-reduce parser
- 两个模型:
  - **Arc-standard:** Left-Reduce, Right-Reduce, Shift
  - **Arc-eager:** Left-Arc, Right-Arc, Reduce, Shift(只介绍)



Transition-based

Nivre’ s parser (Nivre, 2003)

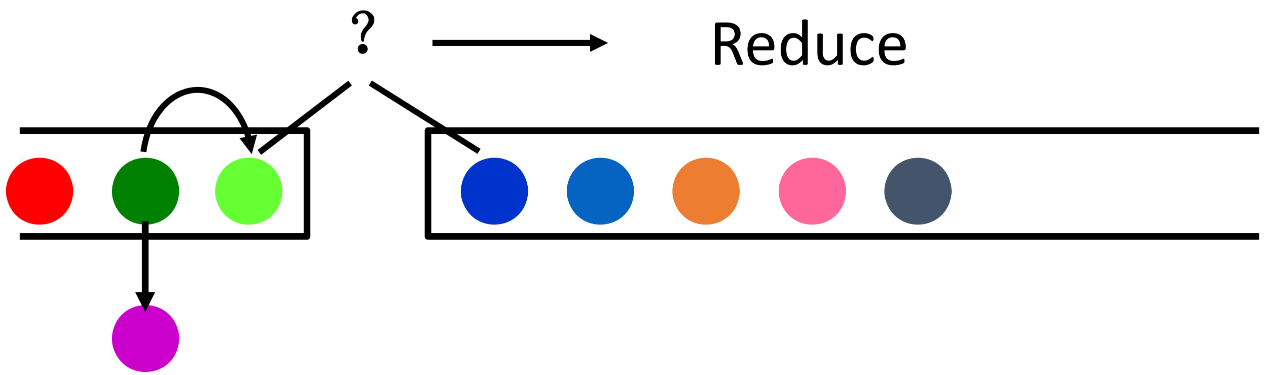
- One-pass shift-reduce parser
- 两个模型:
  - **Arc-standard:** Left-Reduce, Right-Reduce, Shift
  - **Arc-eager:** Left-Arc, Right-Arc, Reduce, Shift(只介绍)



Transition-based

Nivre’ s parser (Nivre, 2003)

- One-pass shift-reduce parser
- 两个模型:
  - **Arc-standard:** Left-Reduce, Right-Reduce, Shift
  - **Arc-eager:** Left-Arc, Right-Arc, Reduce, Shift(只介绍)

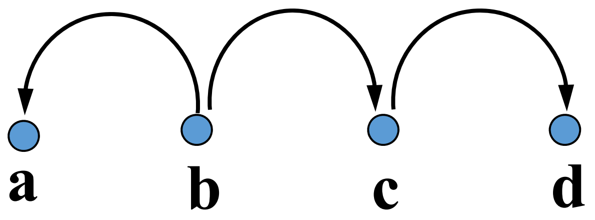


Transition-based

- Nivre’ s parser (Nivre, 2003)
  - One-pass shift-reduce parser
  - 两个模型:
    - **Arc-standard:** Left-Reduce, Right-Reduce, Shift
    - **Arc-eager:** Left-Arc, Right-Arc, Reduce, Shift(只介绍)

Reduce后

Arc-eager：有关系就标，要不要**reduce**单独判断



[a b]→Left-Arc→Shift→Right-Arc→Right-Arc→Reduce→Reduce

初始化: (nil, I, Ø)

结束: (S, nil, A)

Arc-eager关注栈顶的token和输入队列的第一个token

Left-Arc$_l$

$$\frac{[\ldots,w_i]_S \quad [w_j,\ldots]_I \quad \neg\exists\, w_k \rightarrow w_i \in A}{[\ldots]_S \quad [w_j,\ldots]_I \quad A \cup \{w_i \xleftarrow{l} w_j\}}$$

Right-Arc$_l$

$$\frac{[\ldots,w_i]_S \quad [w_j,\ldots]_I}{[\ldots,w_i, w_j]_S \quad [\ldots]_I \quad A \cup \{w_i \xrightarrow{l} w_j\}}$$

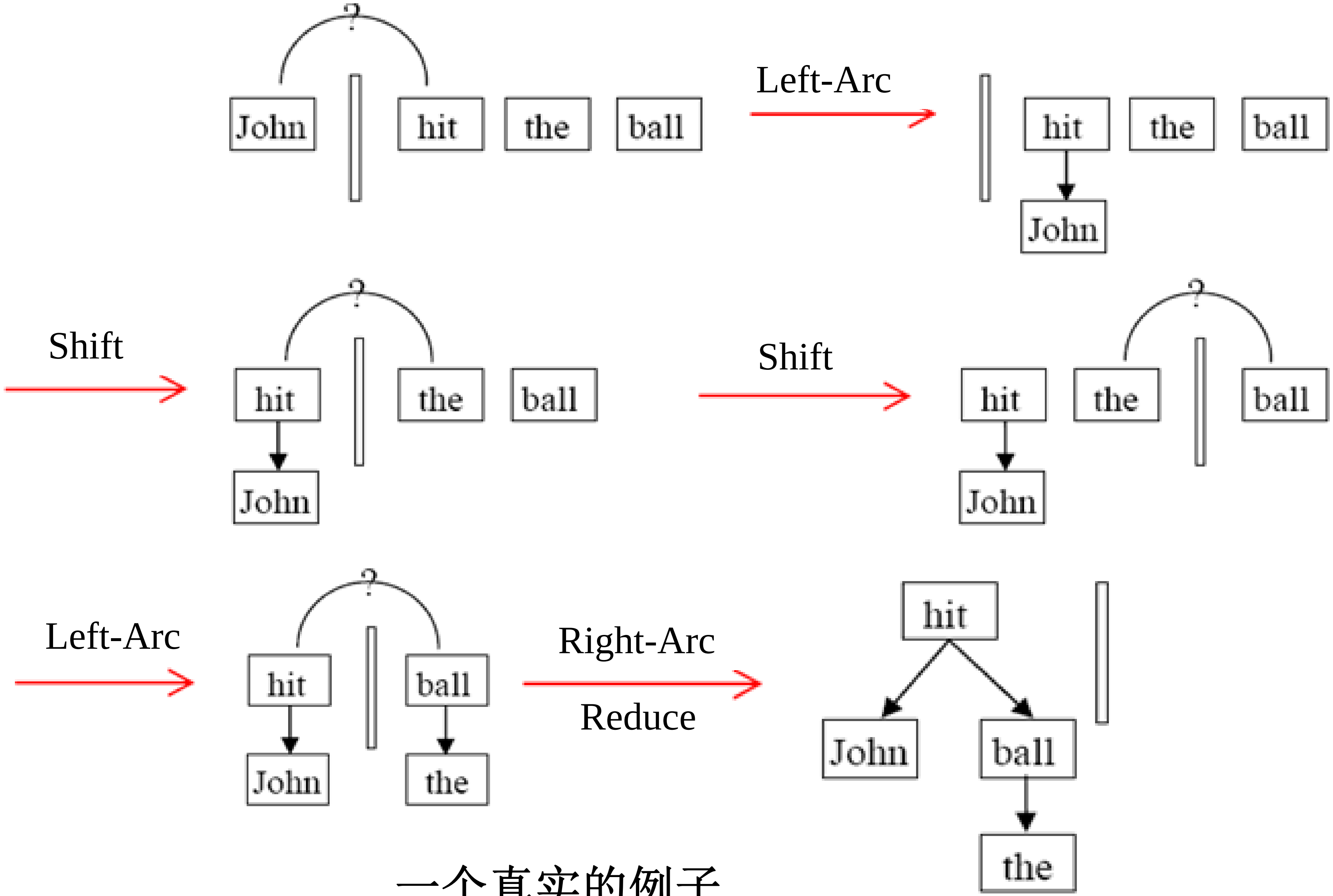
Reduce

$$\frac{[\ldots,w_i]_S \quad [\ldots]_I \quad \exists\, w_k \rightarrow w_i \in A}{[\ldots]_S \quad [\ldots]_I}$$

Shift

$$\frac{[\ldots]_S \quad [w_i,\ldots]_I}{[\ldots,w_i]_S \quad [\ldots]_I}$$

Actions in Arc-eager model


John
hit
the
ball
Left-Arc
hit
the
ball
John
Shift
hit
the
ball
John
Shift
hit
the
ball
John
Left-Arc
hit
ball
John
the
Right-Arc
Reduce
hit
John
ball
the

一个真实的例子

Transition-based

Transition-based方法

- 预测出最优的动作序列，就可以得到依存树
- 通常看作一个Sequential classifications问题：
  - 在每一个当前状态下预测一个最优动作，整体上是一个分类序列
  - 因此，是一个确定性模型
  - 可以采用任一种分类器

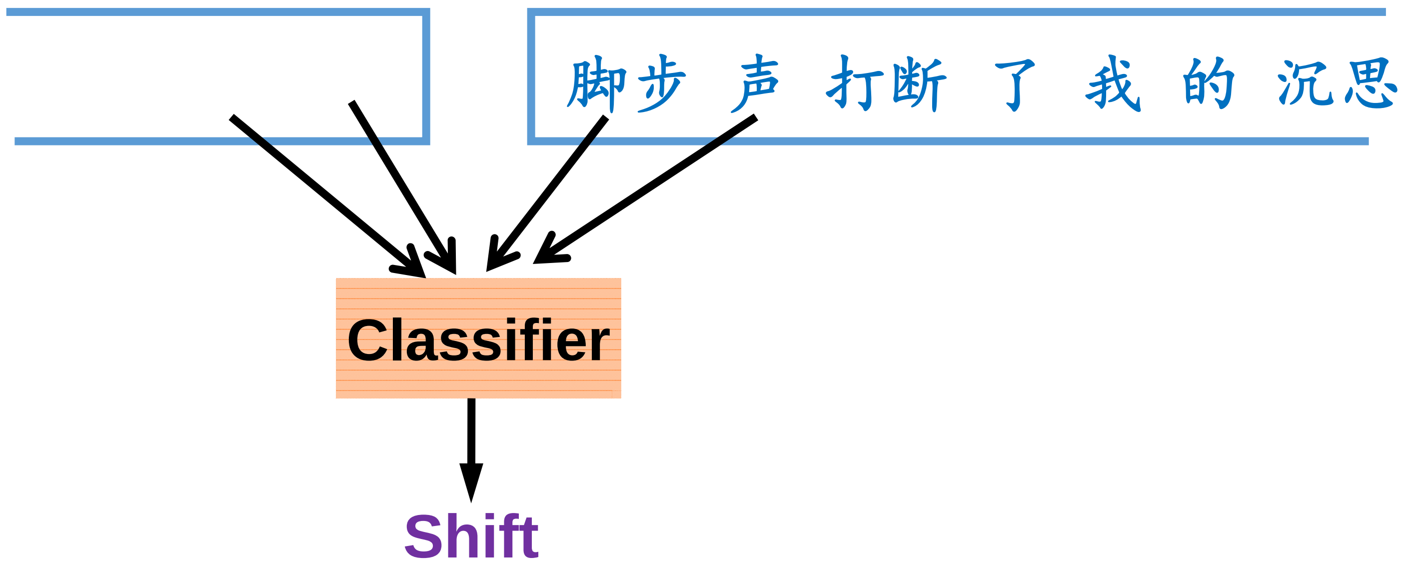
Transition-based

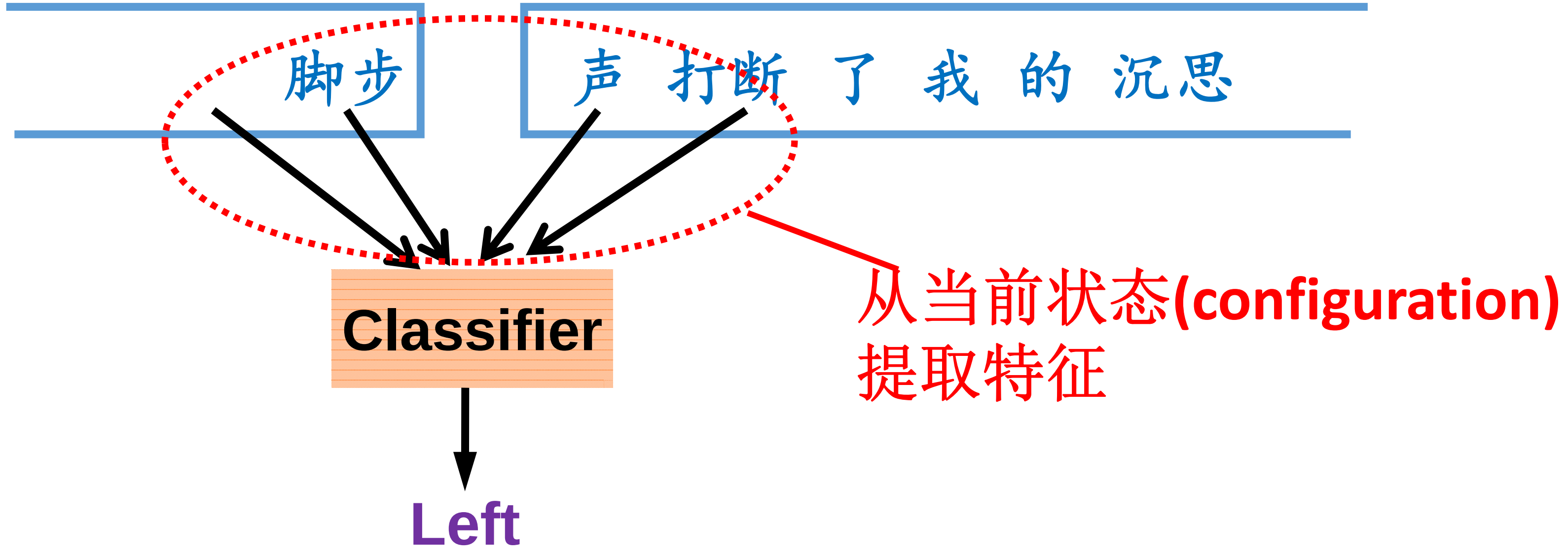
◻ Transition-based方法优缺点

- 通常是一个确定性模型
- Greedy，不是全局寻优，时间复杂度$O(n^2)$或$O(n)$
- 易使用动态特征
- 可以方便预测relation type
- 存在错误传递问题，思考如何缓解？

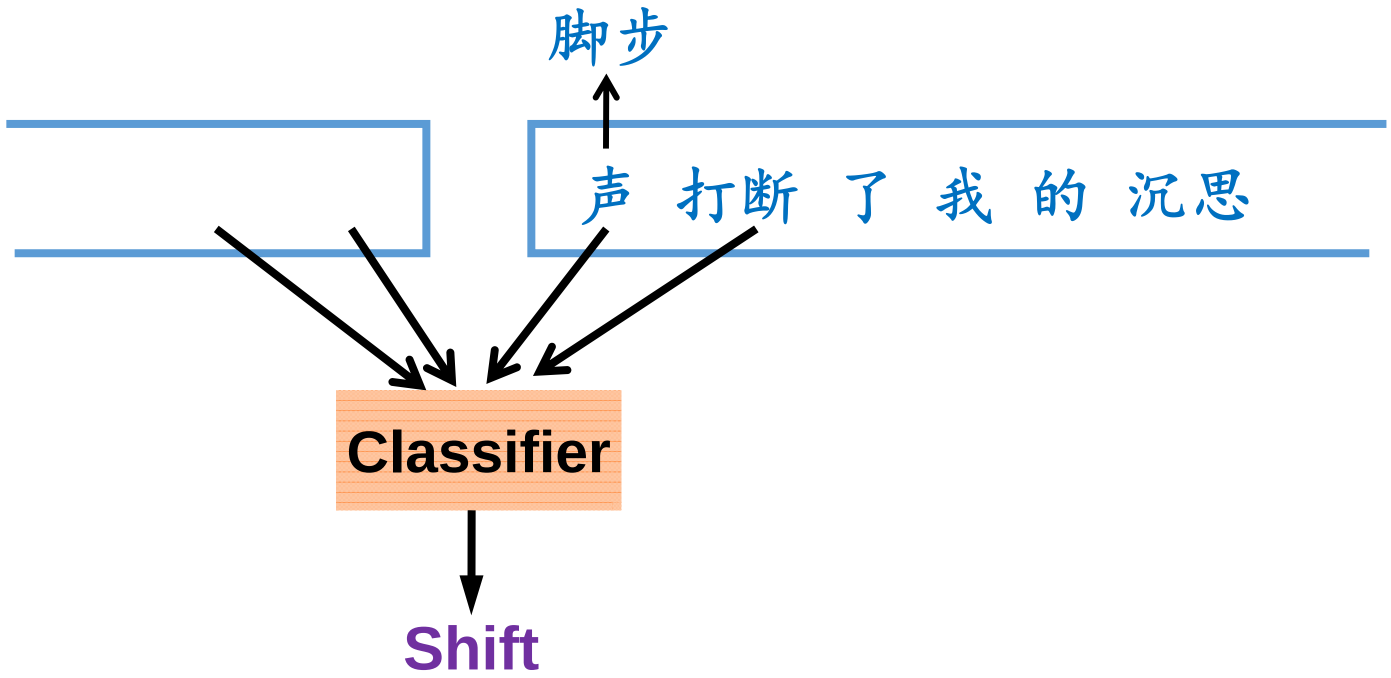
Transition-based

➢ 一个arc-eager parser举例

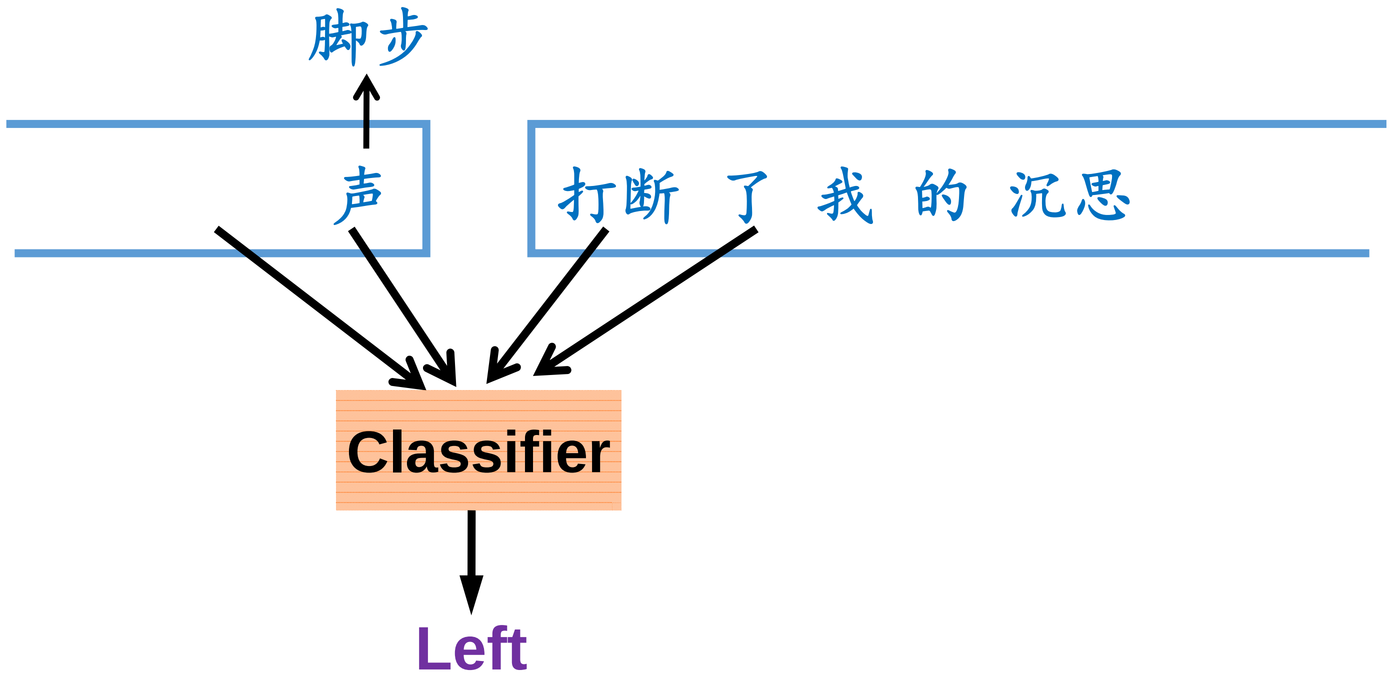




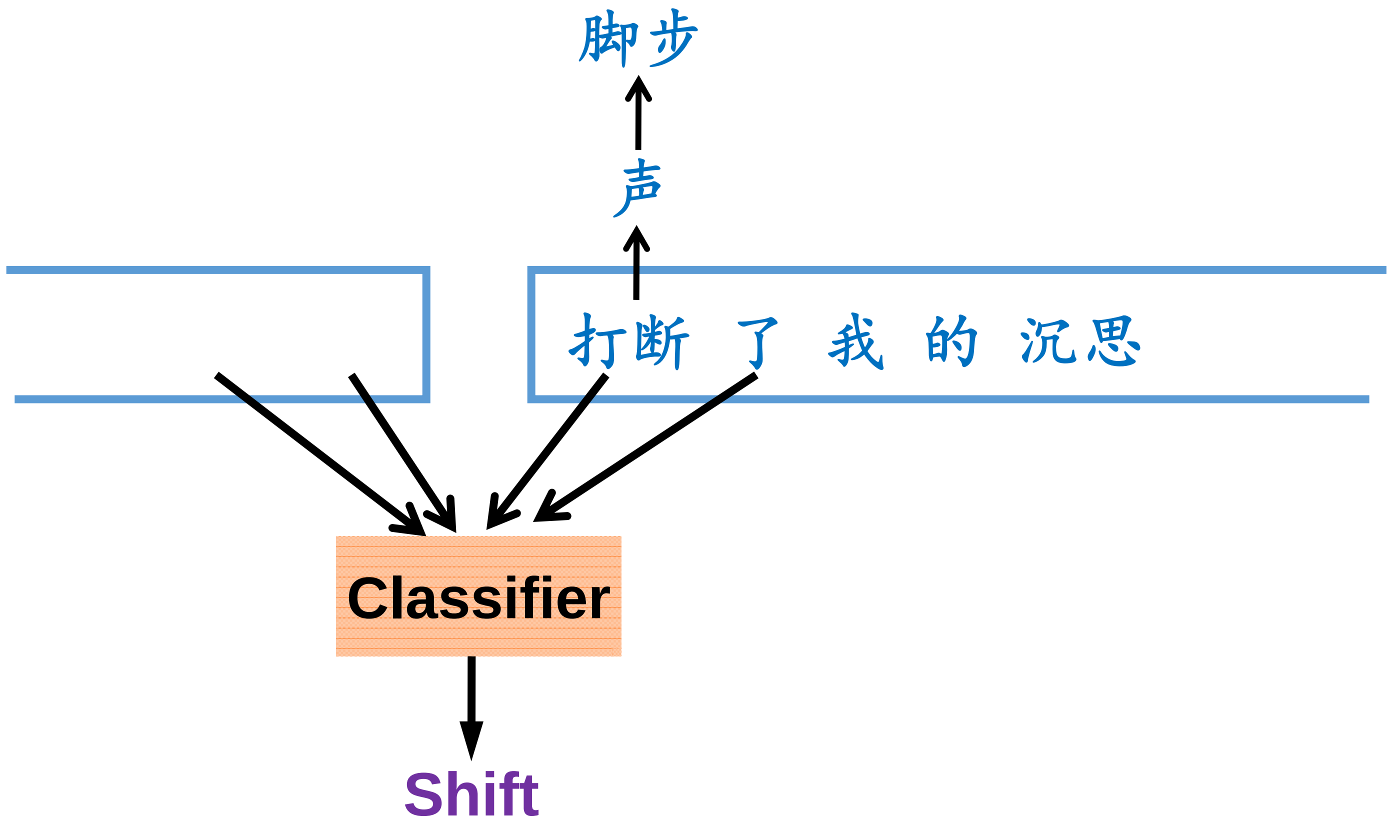
Shift



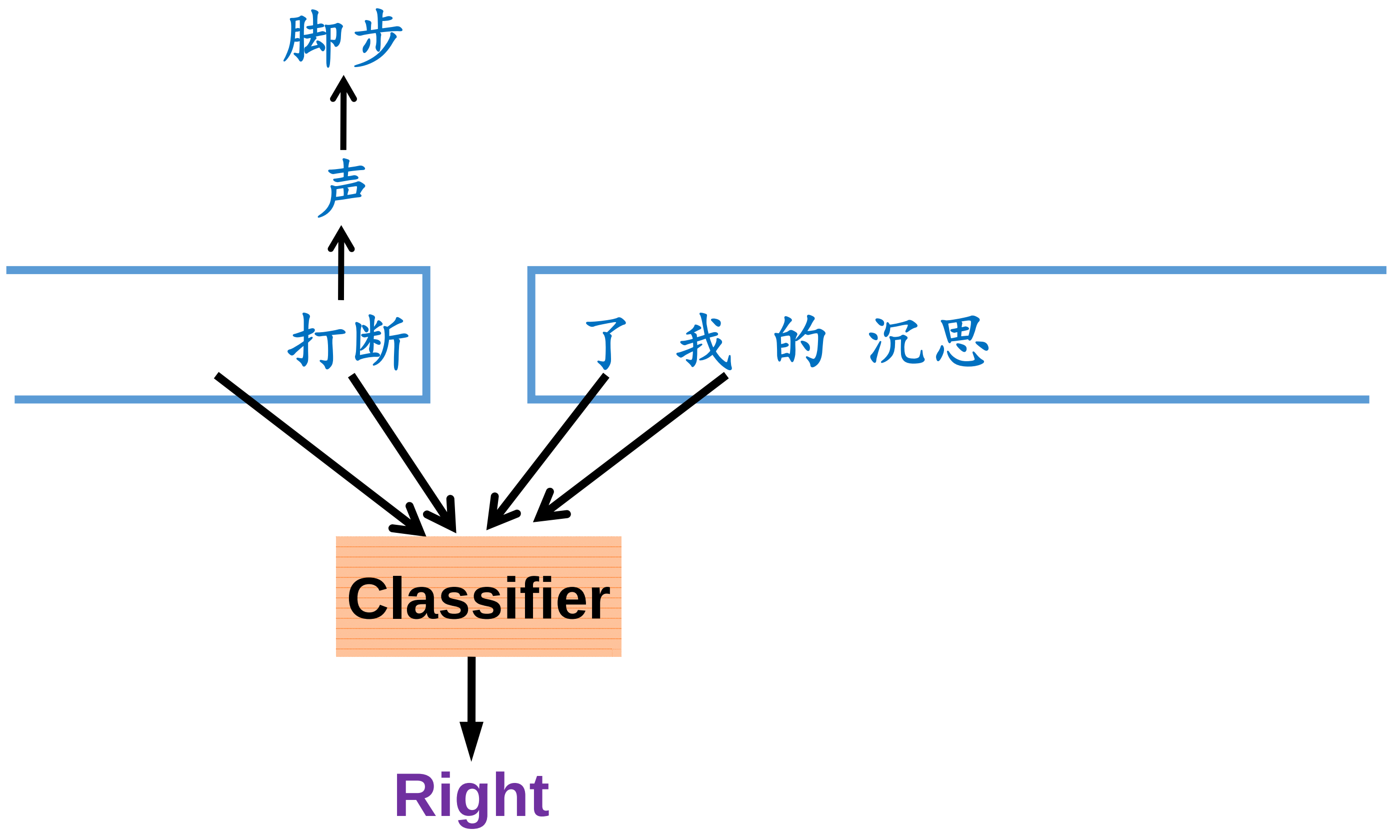
Shift Left



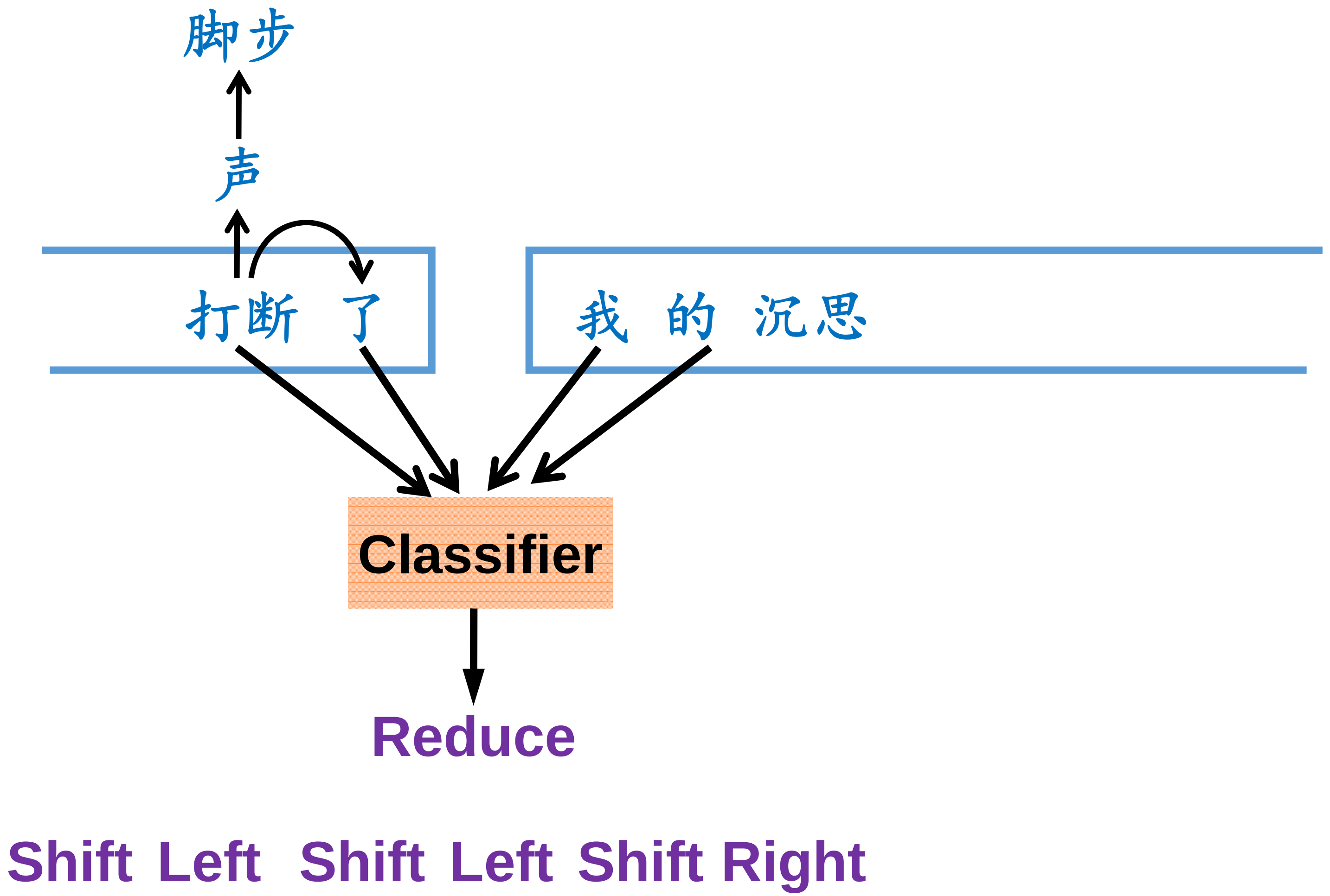
Shift Left Shift



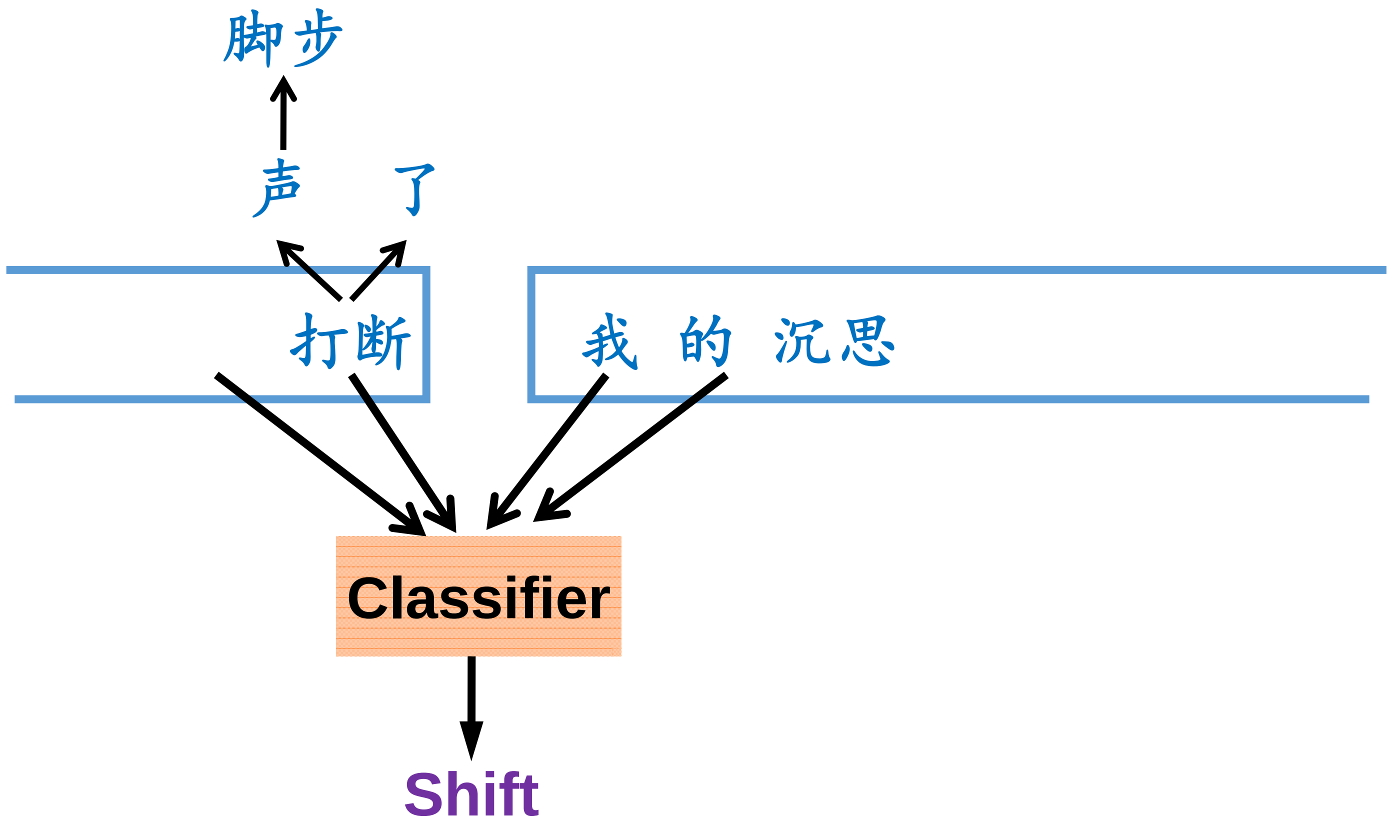
Shift Left Shift Left



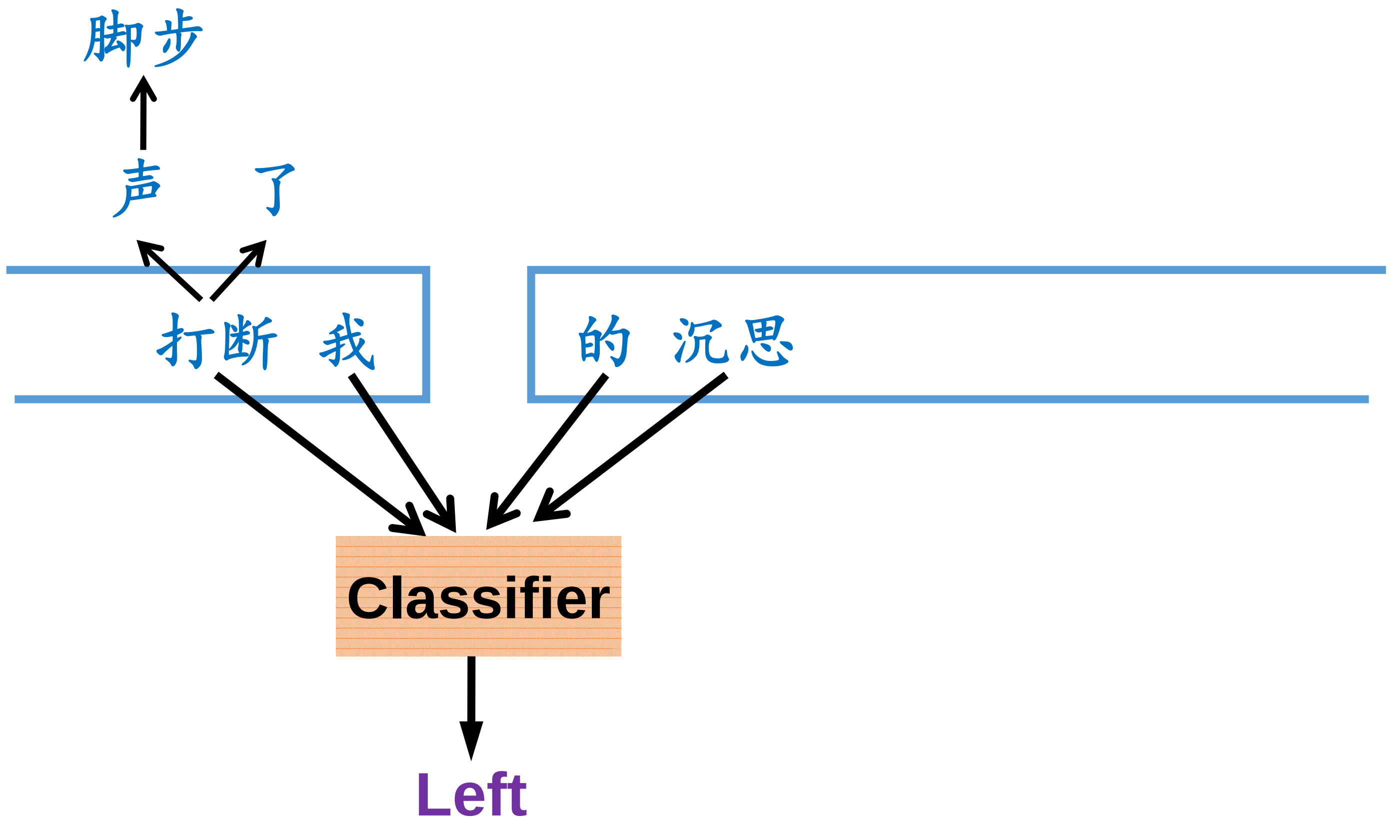
Shift Left Shift Left Shift



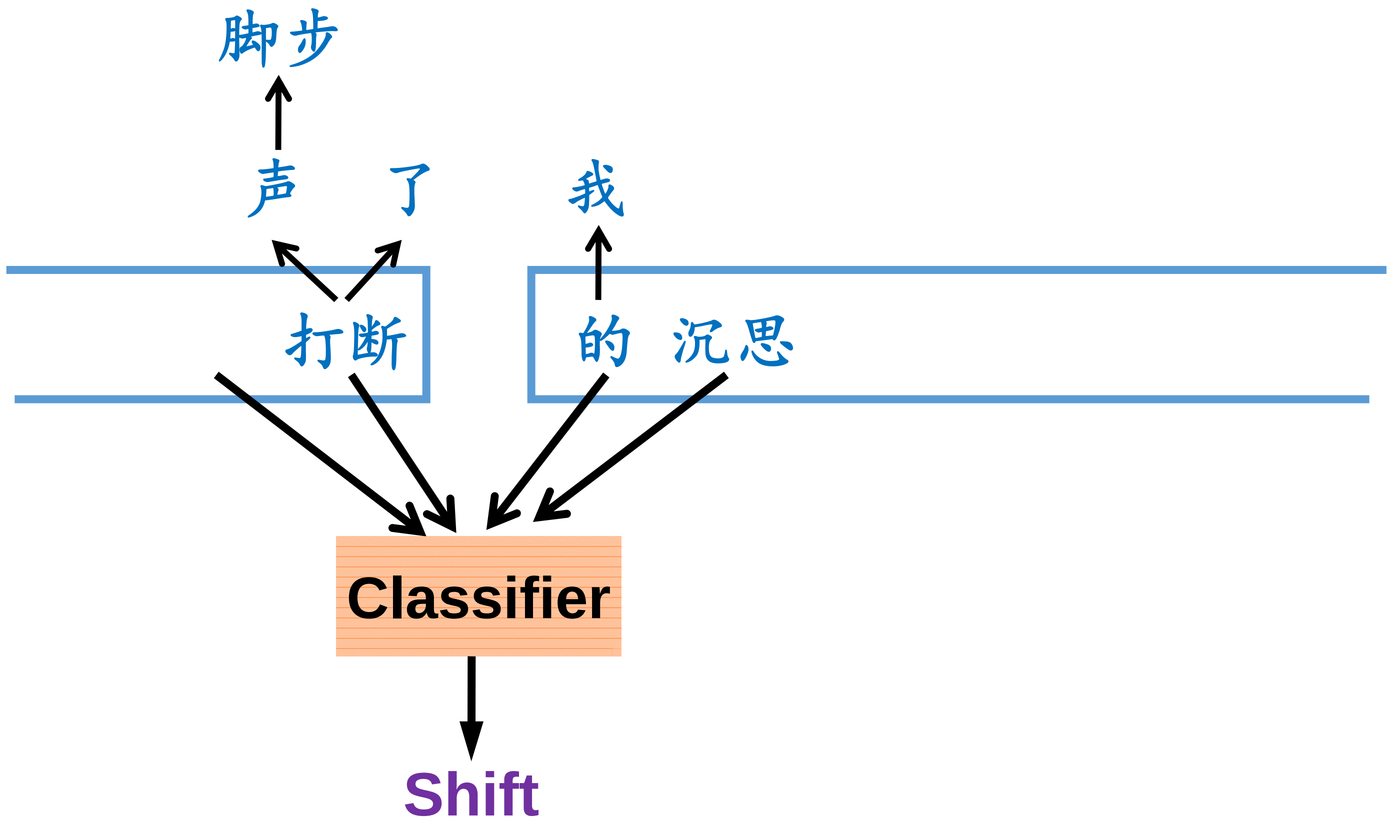
脚步
声
打断 了
我 的 沉思
Classifier
Reduce
Shift Left Shift Left Shift Right



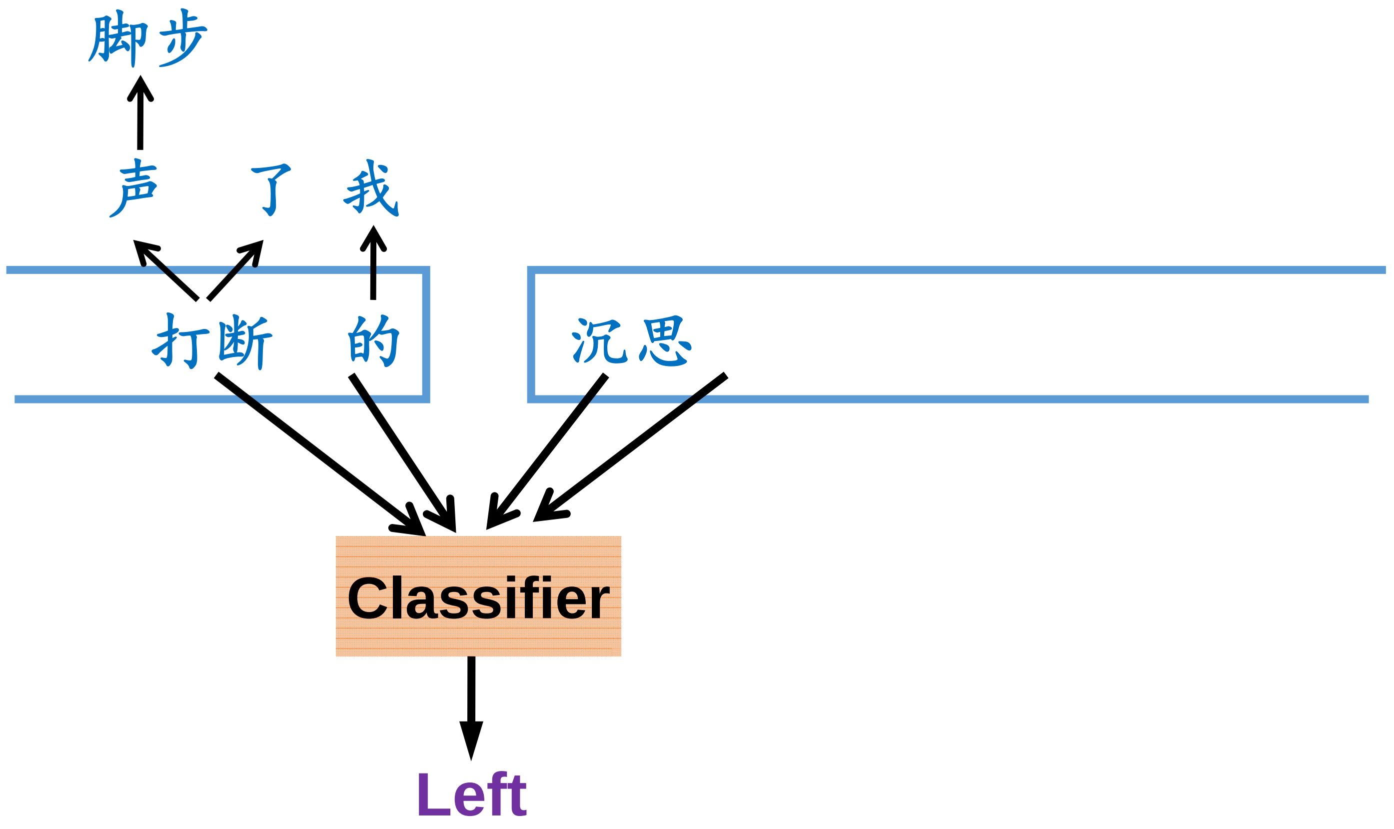
Shift Left Shift Left Shift Right Reduce



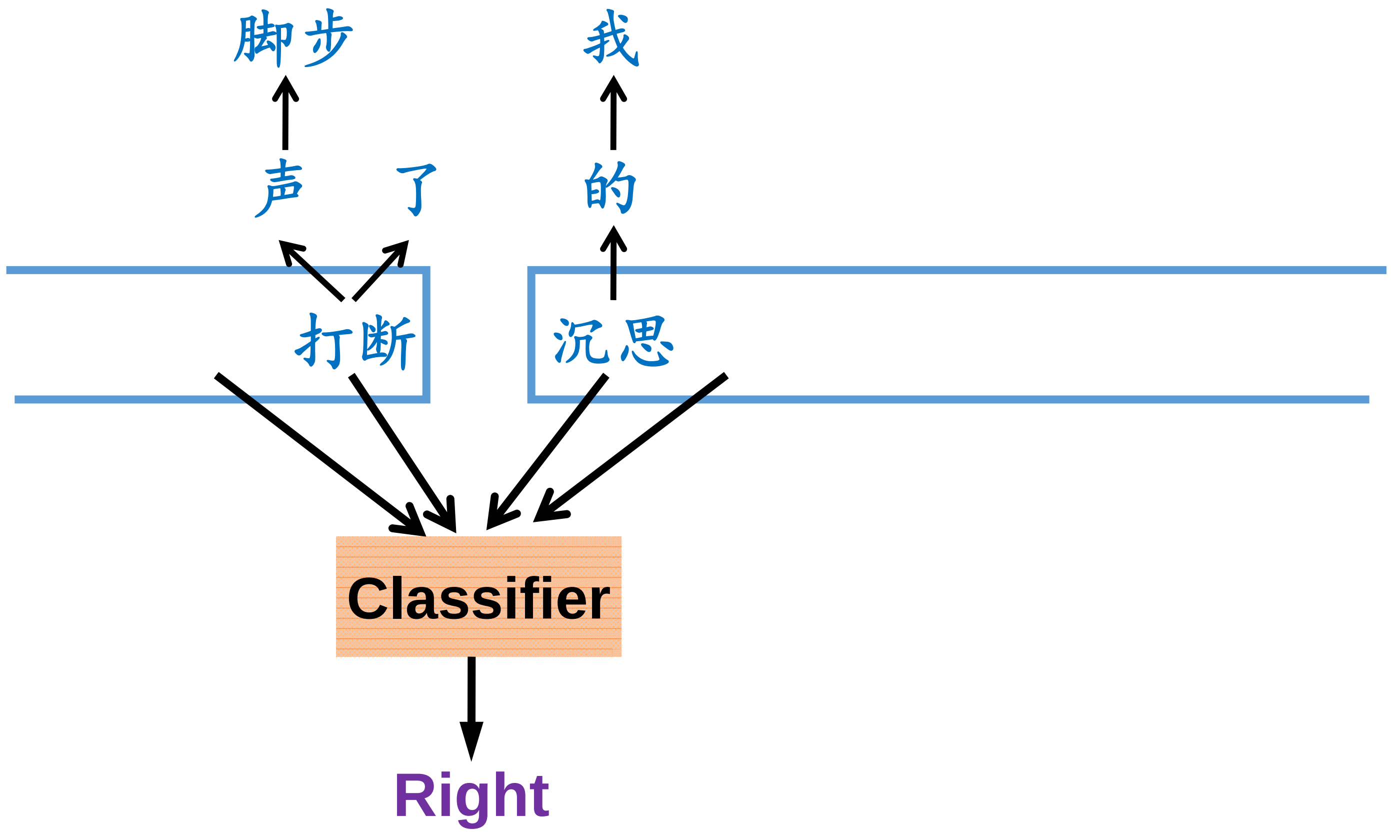
Shift Left Shift Left Shift Right Reduce Shift



Shift Left Shift Left Shift Right Reduce Shift Left

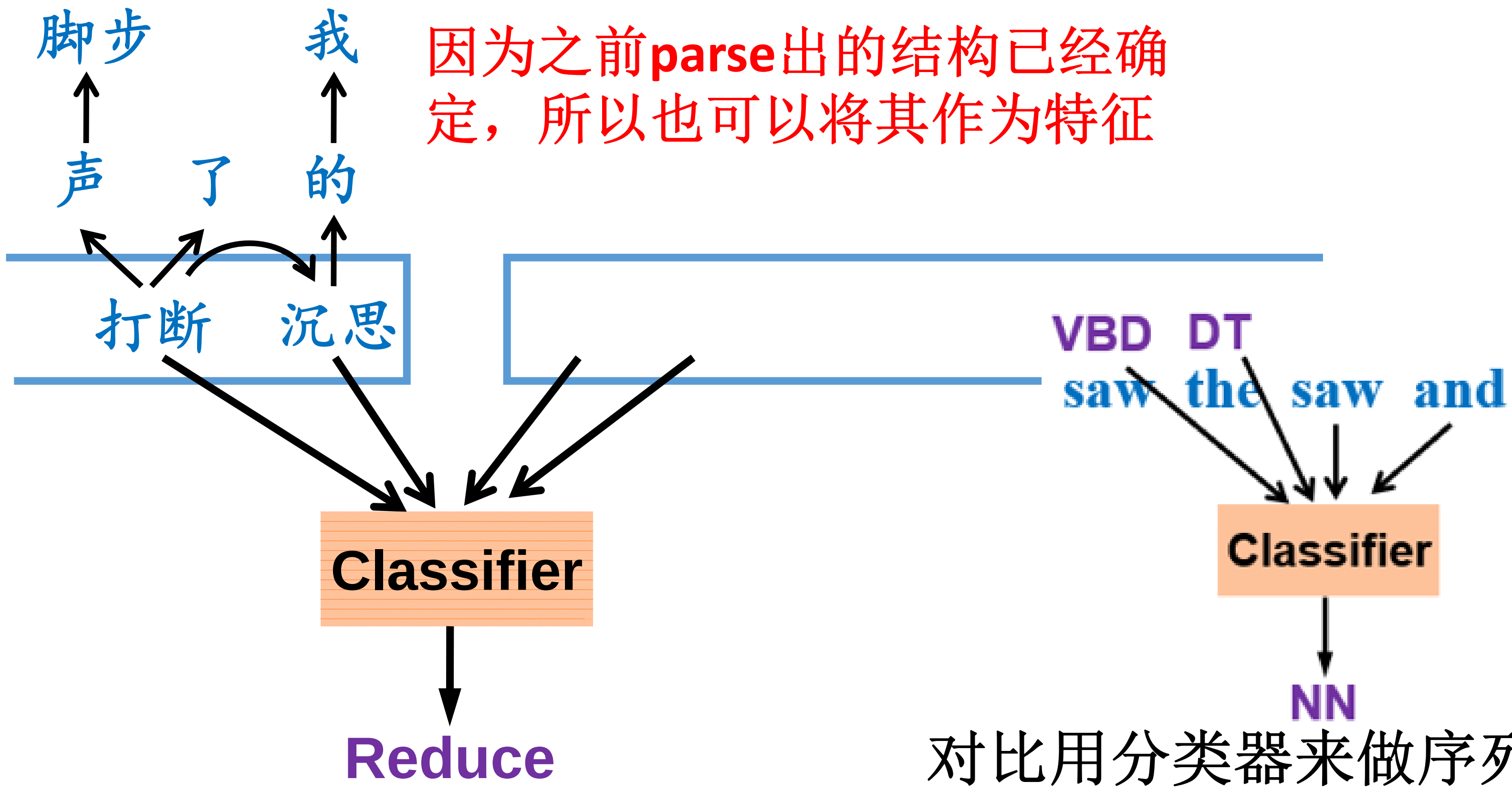


Shift Left Shift Left Shift Right Reduce Shift Left Shift



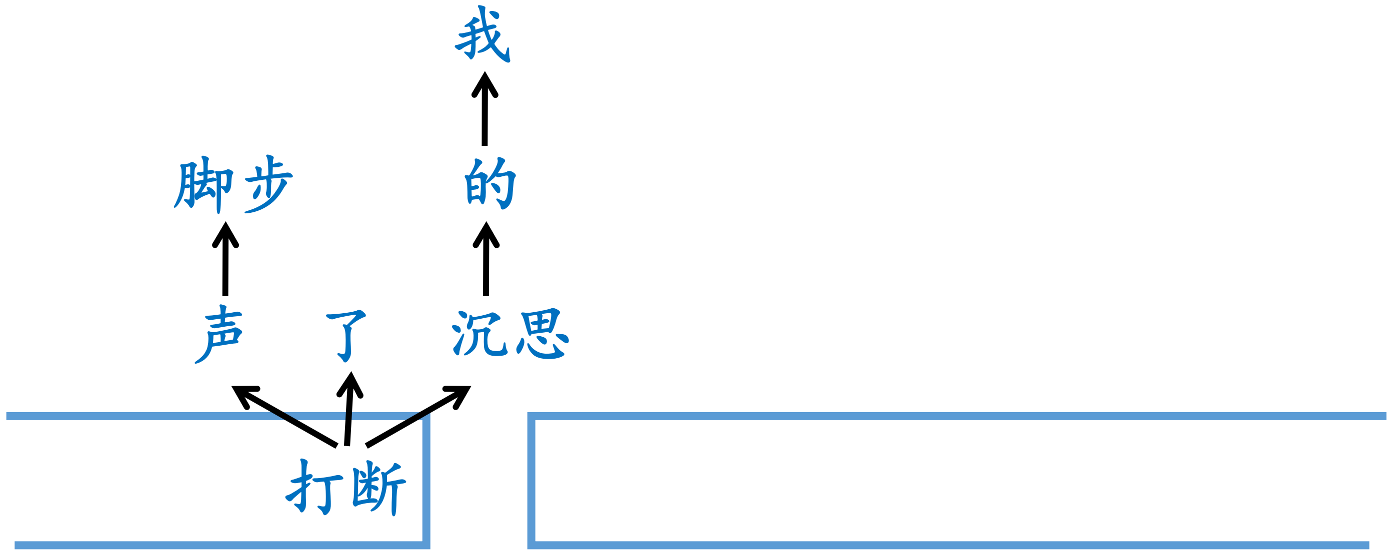
Shift Left Shift Left Shift Right Reduce Shift Left Shift Left

因为之前parse出的结构已经确定，所以也可以将其作为特征



对比用分类器来做序列标注

Shift Left Shift Left Shift Right Reduce Shift Left Shift Left Right



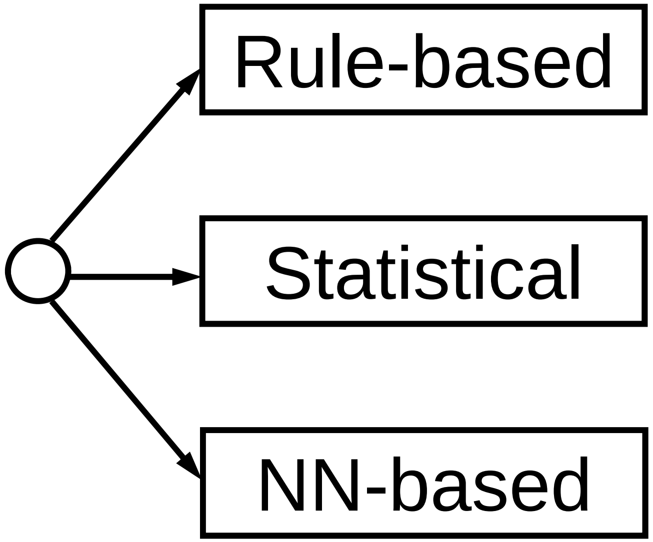
✓ 这个过程和序列标注里用多个分类器来预测每一个位置的词性是类似的，都是确定性的方法

✓ 把一个结构预测(结构化)问题转变为一个序列预测问题(思考为什么不是一个序列标注问题？)

Shift Left Shift Left Shift Right Reduce Shift Left Shift Left Right Reduce

依存结构句法分析

依存分析方法



D-rules

Graph-based, transition-based

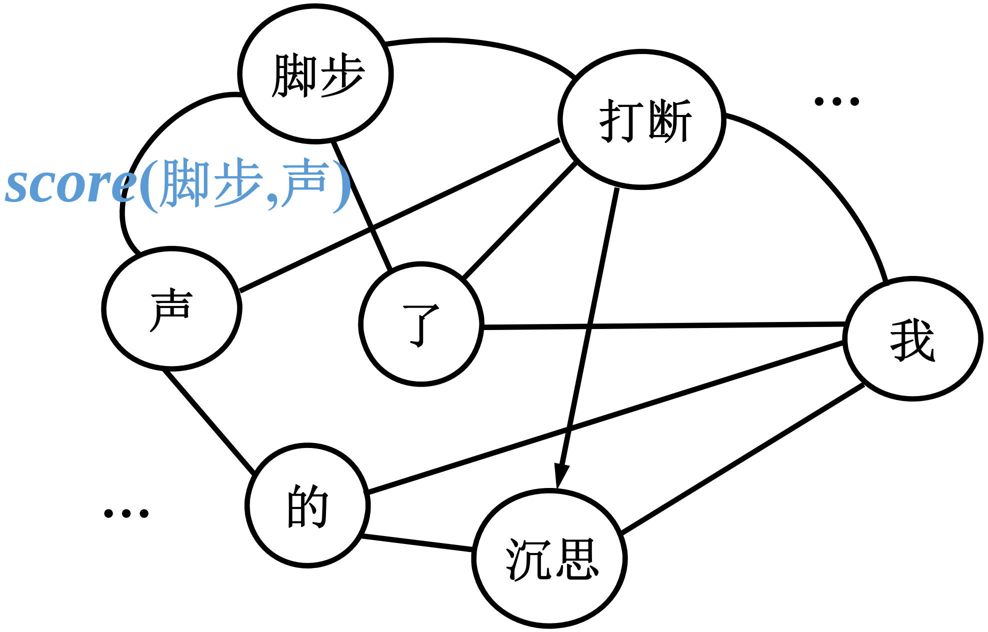
Graph-based, transition-based

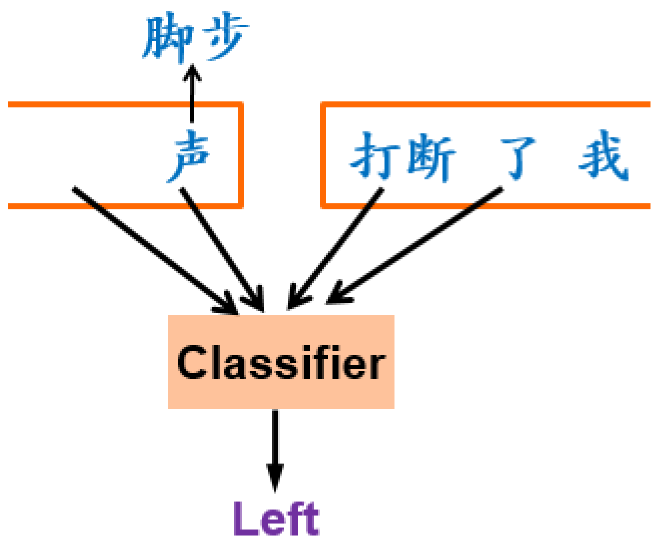
依存分析的NN方法

- 经典的概率依存句法分析模型怎样演化到神经网络模型？
  1. 只是特征用分布式表示，算法不变
  2. 新的计算机制

依存分析的NN方法

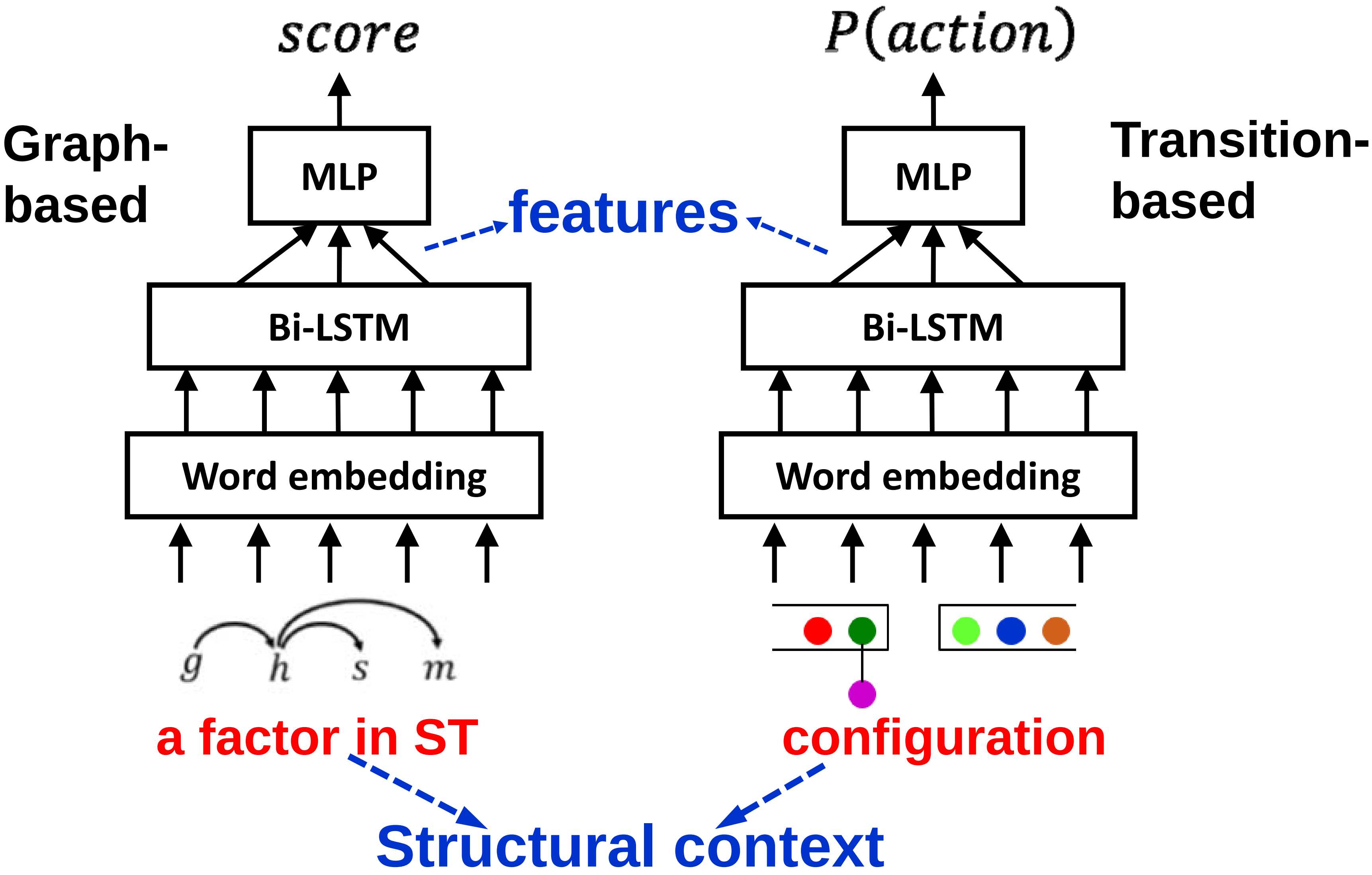
- 回顾两种依存分析方法





$$s(x,y)=\sum_{(i,j\in y)} score(i,j) \quad t^* = \operatorname{argmax} \sum_{(t\in T)} score(c,t)$$

1. 只是特征用分布式表示，算法不变



ACL2016： “ Graph-based Dependency Parsing with Bidirectional LSTM. ”

一个基于特征的**graph-based**模型

ACL2016：“Graph Bidirectional LSTM.

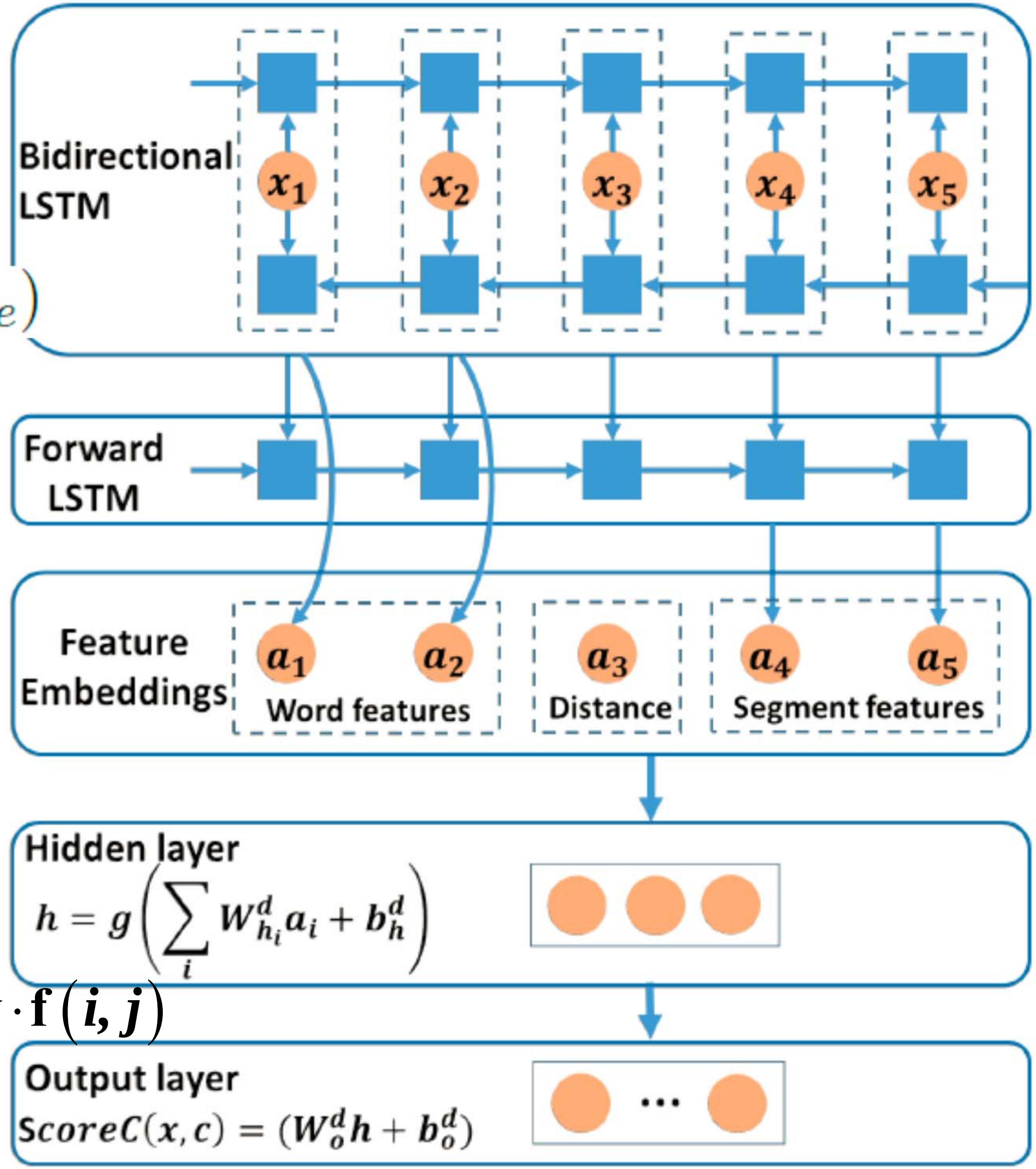
$$x_i = g(w_e[e_{w_i}; e_{p_i}] + b_e)$$

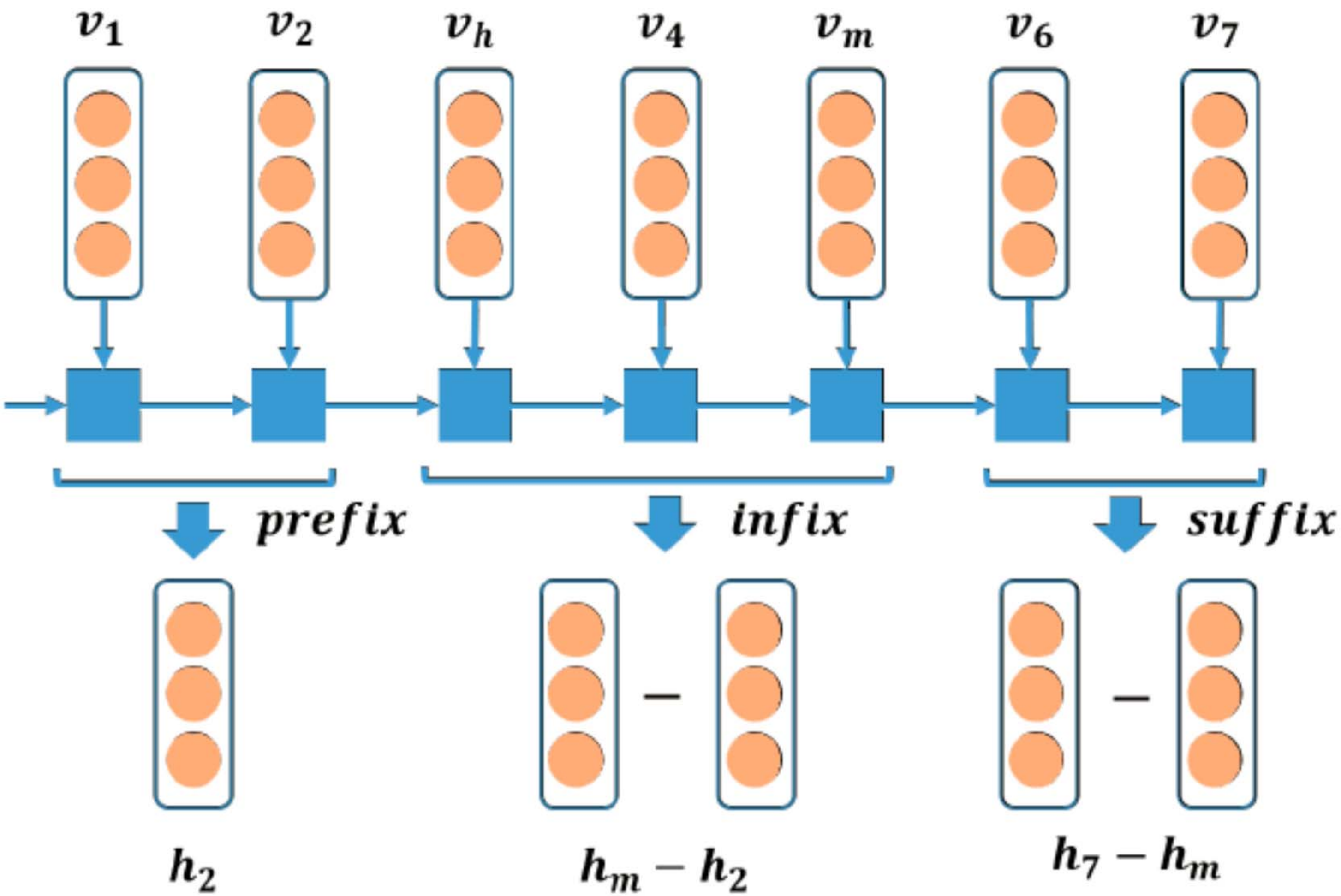
$$v_t = \overrightarrow{h}_t + \overleftarrow{h}_t$$

一个基于特征的模型

$$ScoreC(x,c) = W_o^d h + b_o^d$$

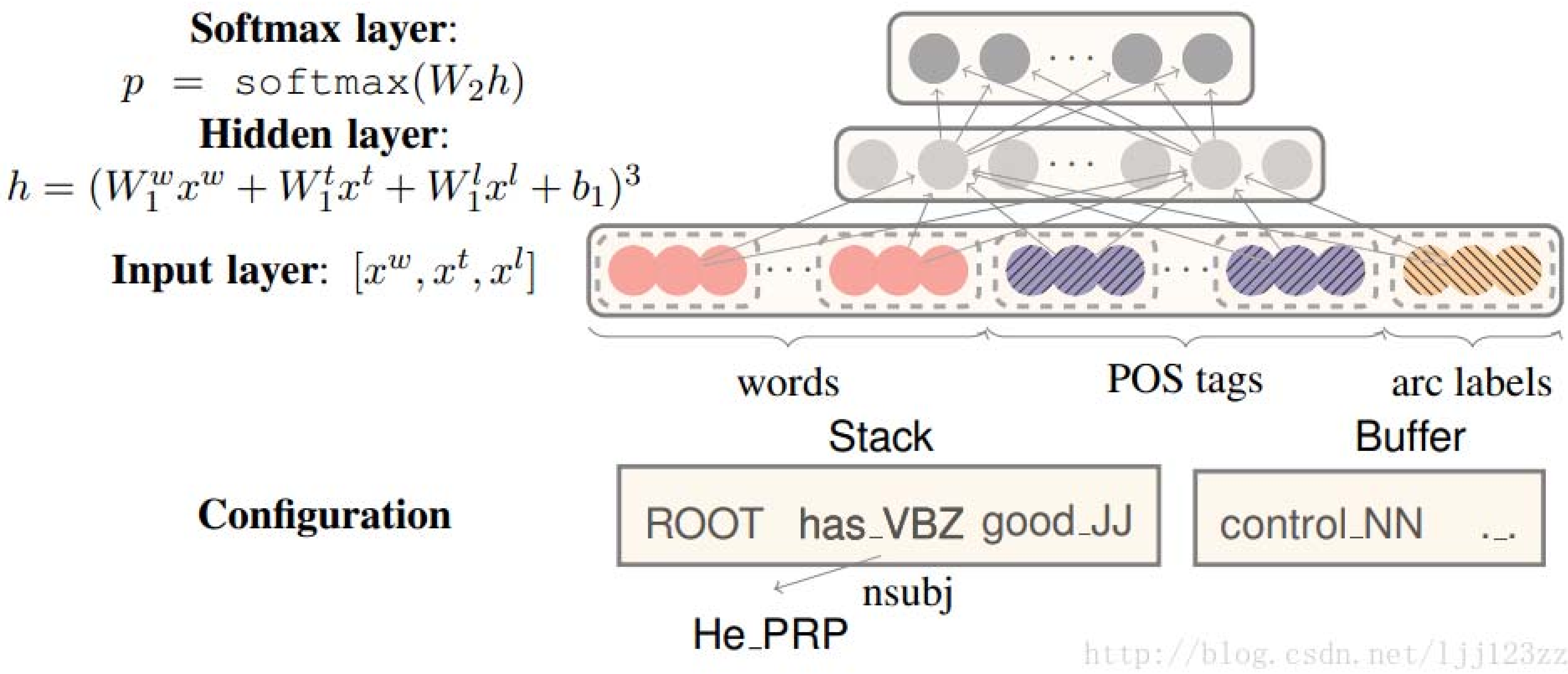
$$s(x,y) = \sum_{(i,j)\in y} s(i,j) = \sum_{(i,j)\in y} \mathbf{w} \cdot \mathbf{f}(i,j)$$





获取段特征(段的分布式表示)

EMNLP2014: “A Fast and Accurate Dependency Parser using Neural Networks. ”



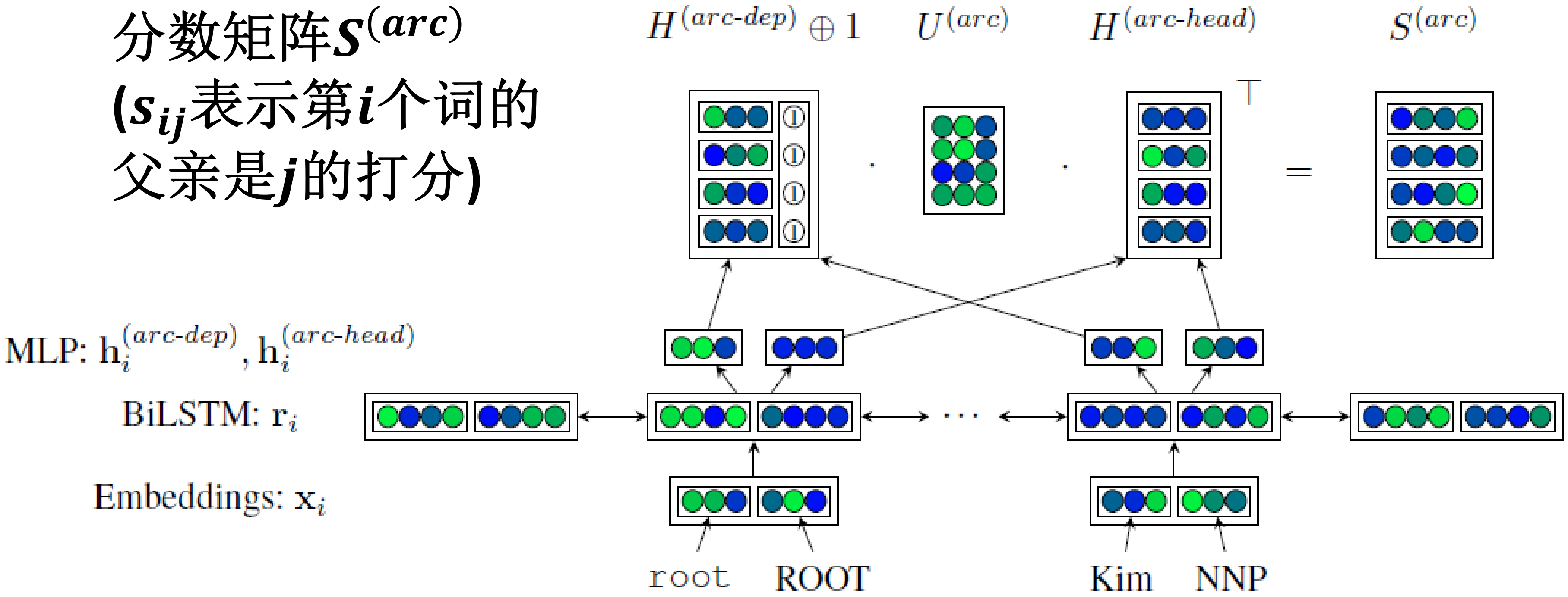
一个基于特征的**transition-based**模型

2. 新的计算机制

➢ 双仿射注意力的graph-based NN模型

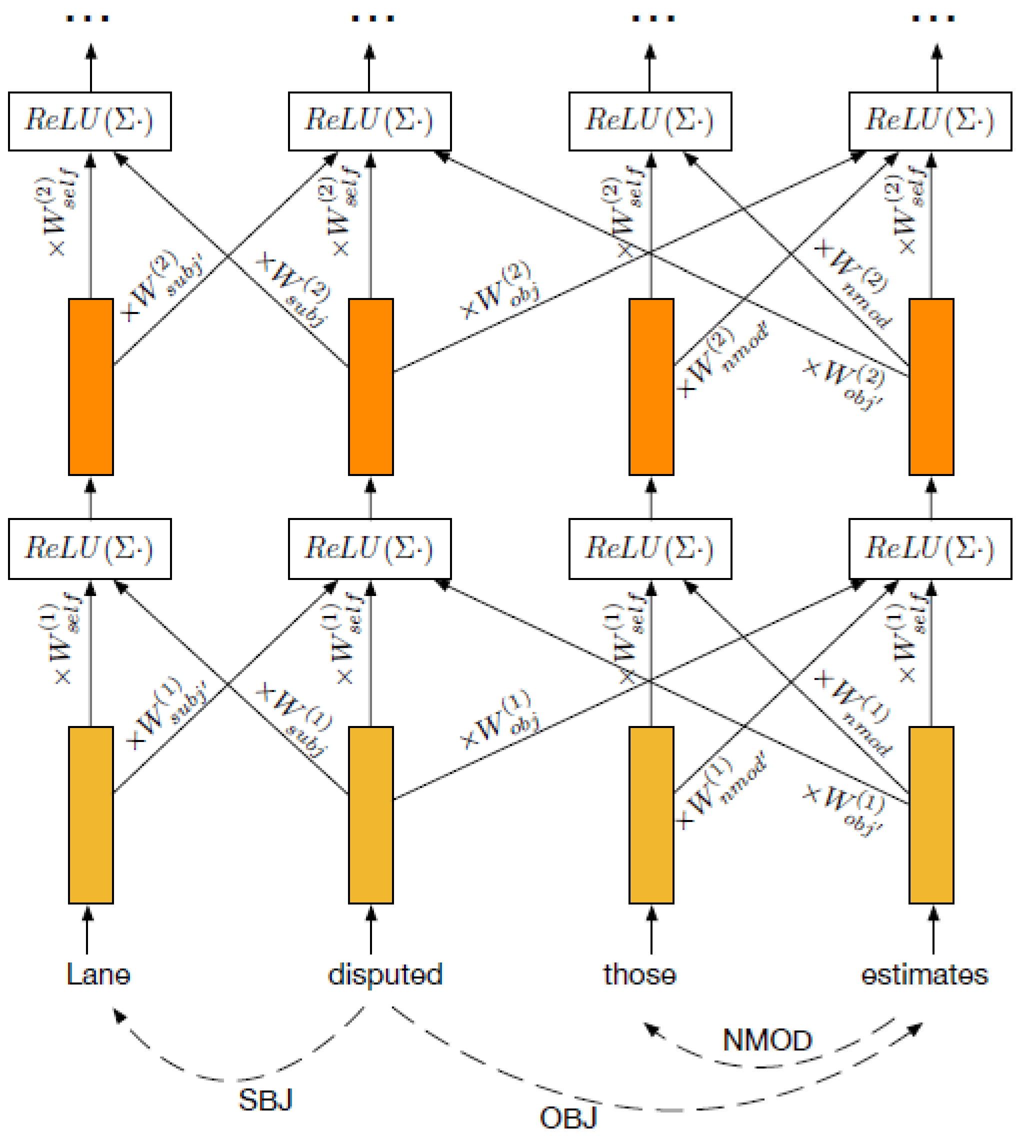
经过一个中间矩阵$\boldsymbol{U}^{(arc)}$仿射变换后，每个**token**以**dep**的身份与以**head**身份的每个**token**进行一次点积，得到**arc**成立的分数矩阵$\boldsymbol{S}^{(arc)}$

($\boldsymbol{s_{ij}}$表示第$\boldsymbol{i}$个词的父亲是$\boldsymbol{j}$的打分**)**

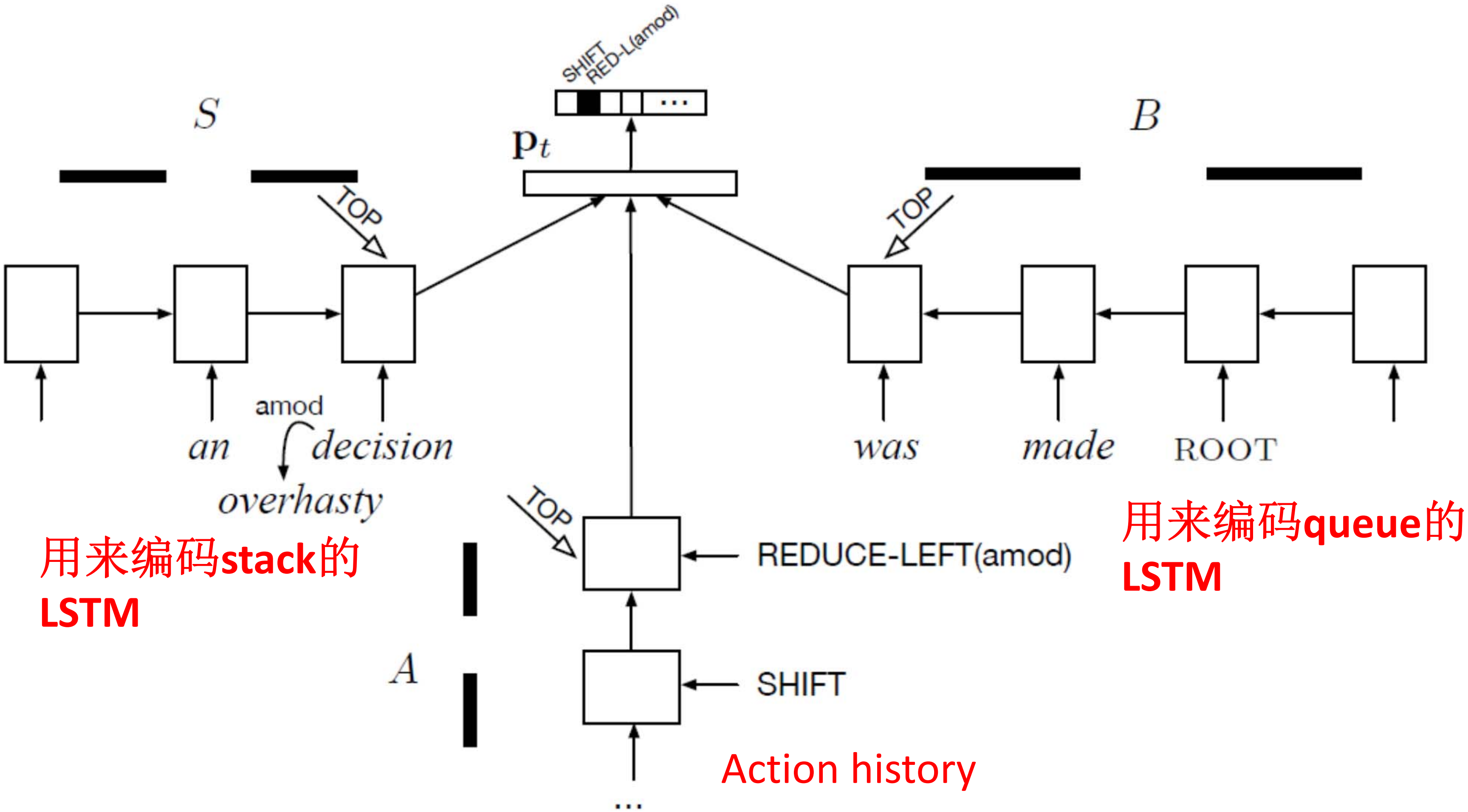


ICLR2017: “Deep Biaffine attention for neural dependency parsing. ”

2017年曾有工作将GCN用于SRL和MT，利用的是相同的思想，但目的是将句法信息融入SRL或MT

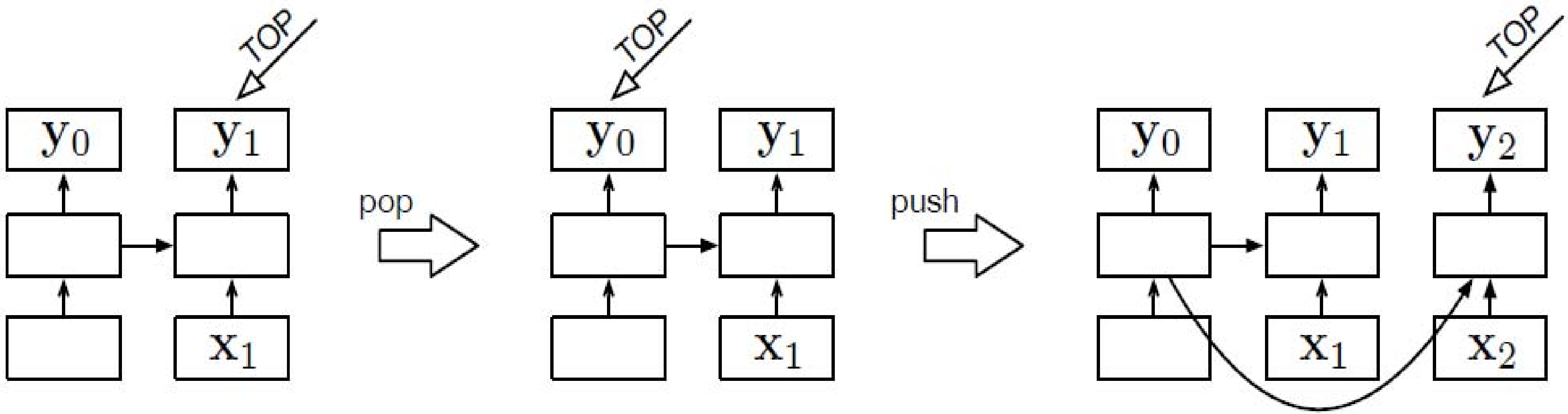


2. 新的计算机制



ACL2015: “Transition-Based Dependency Parsing with Stack Long Short-Term Memory. ”

2. 新的计算机制



你需要一个“新的”LSTM来编码栈—— a stack LSTM

ACL2015: “Transition-Based Dependency Parsing with Stack Long Short-Term Memory. ”

依存句法分析

- 语料(自行从短语结构语料转换)
  - 英语：PTB；汉语：CTB
- Parser
  - https://nlp.stanford.edu/software/
  - Graph-based：MSTParser
    https://sourceforge.net/projects/mstparser/
  - Transition-based：MaltParser
    http://maltparser.org/index.html
  - 哈工大 LTP
    http://ltp.ai/index.html

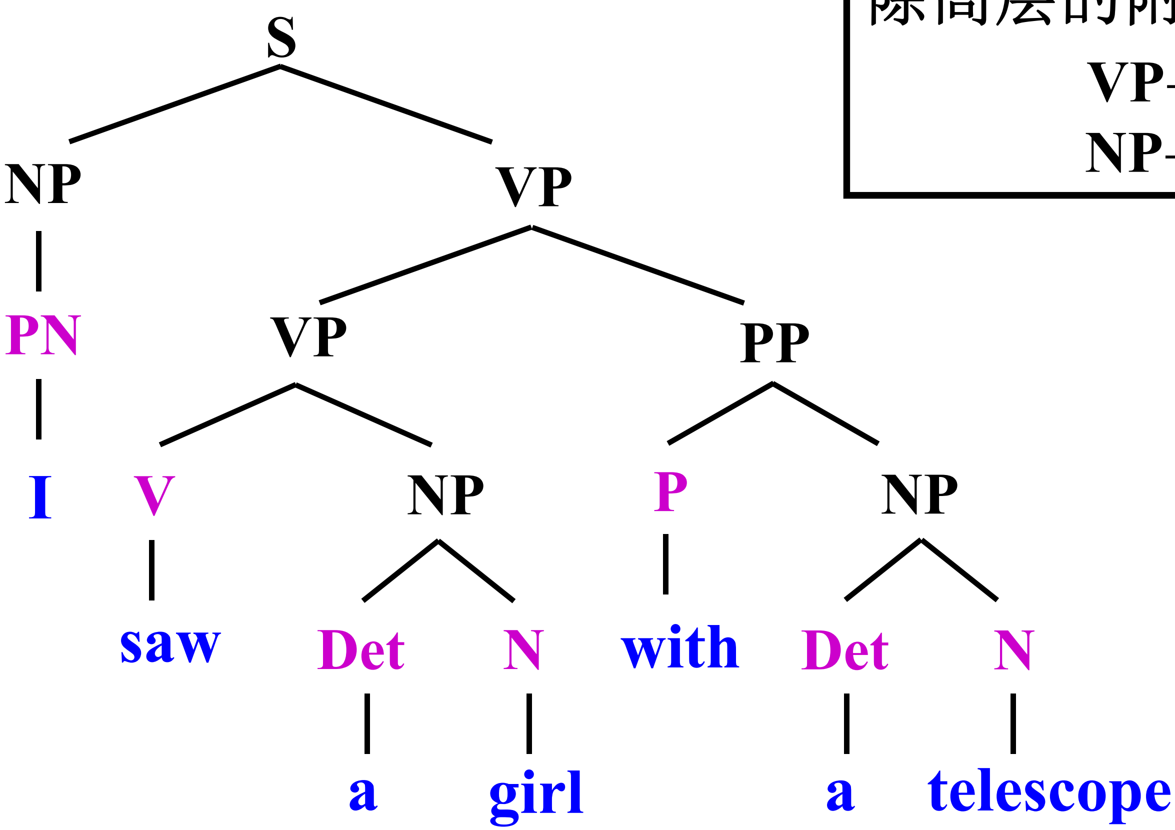
大纲

- 句法分析
  - 短语结构句法分析
  - 依存结构句法分析
  - 浅层句法分析
- 篇章分析
  - 篇章结构分析

浅层句法分析

- shallow parsing，相对于完全句法分析(full parsing)
- 也称为部分句法分析或组块识别(chunking)
- 识别简单的成分，不获取整个句法树
  - 避开了较难处理的结构歧义问题，将其留到下一阶段，期待更多的消歧信息
  - 降低完全句法分析的复杂度
  - 将句法分析独立为两部分，有利于不同分析模型或技术的互补

浅层句法分析



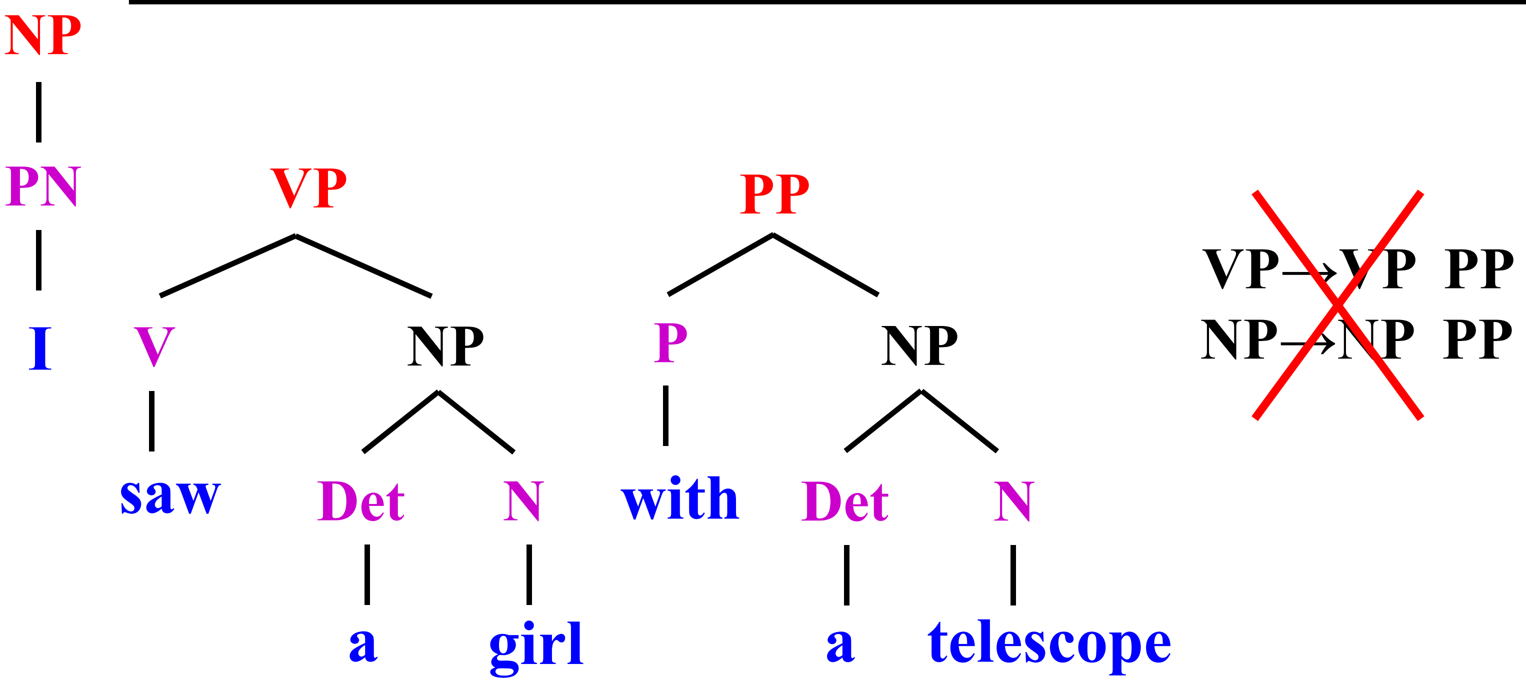
完全句法分析中，需要排除高层的附着歧义：

VP→VP PP

NP→NP PP

浅层句法分析

在浅层句法分析中，附着歧义可以被保存到高级阶段，在高级阶段使用高级的技术和更深入的知识

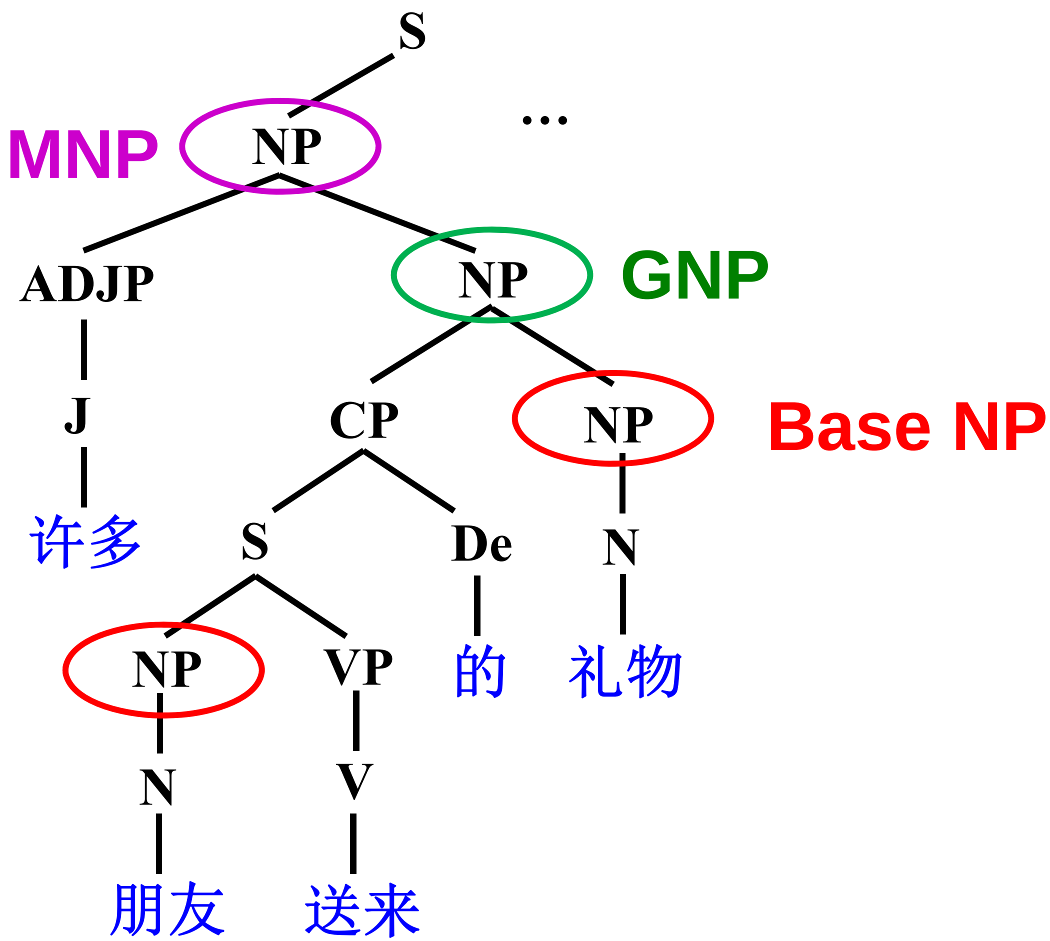


浅层句法分析

◻ 主要任务

- 基本名词短语识别：识别base NP
- 组块识别(chunking)：识别所有基本短语
- 其他不以整个句法树为目标的句法分析任务，如最大短语识别
  - ✓ **Base NP**
  - ✓ **Maximal (length) NP**
  - ✓ **General NP**

浅层句法分析



浅层句法分析

汉语NP的嵌套

阿根廷	NR	BS	O	BS	BS
拉里奥哈省	NR	IH	O	IS	IS
利用	VV	O	O	IS	IS
中国	NR	BS	BS	IS	IS
银行	NN	IH	IS	IS	IS
提供	VV	O	IS	IS	IS
贷款	NN	O	IH	IS	IS
购买	VV	O	O	IS	IS
的	DEC	O	O	IS	IS
中国	NR	O	O	IS	IS
机械	NN	O	O	IS	IS
产品	NN	O	O	IH	IS
的	DEG	O	O	O	IS
交接	NN	O	O	O	IS
仪式	NN	O	O	O	IH
今天	NT	O	O	O	O
在	P	O	O	O	O
该	DT	BS	O	O	O
省	NN	IS	O	O	O
查米卡尔市	NR	IH	O	O	O
举行	VV	O	O	O	O
。	PU	O	O	O	O

浅层句法分析

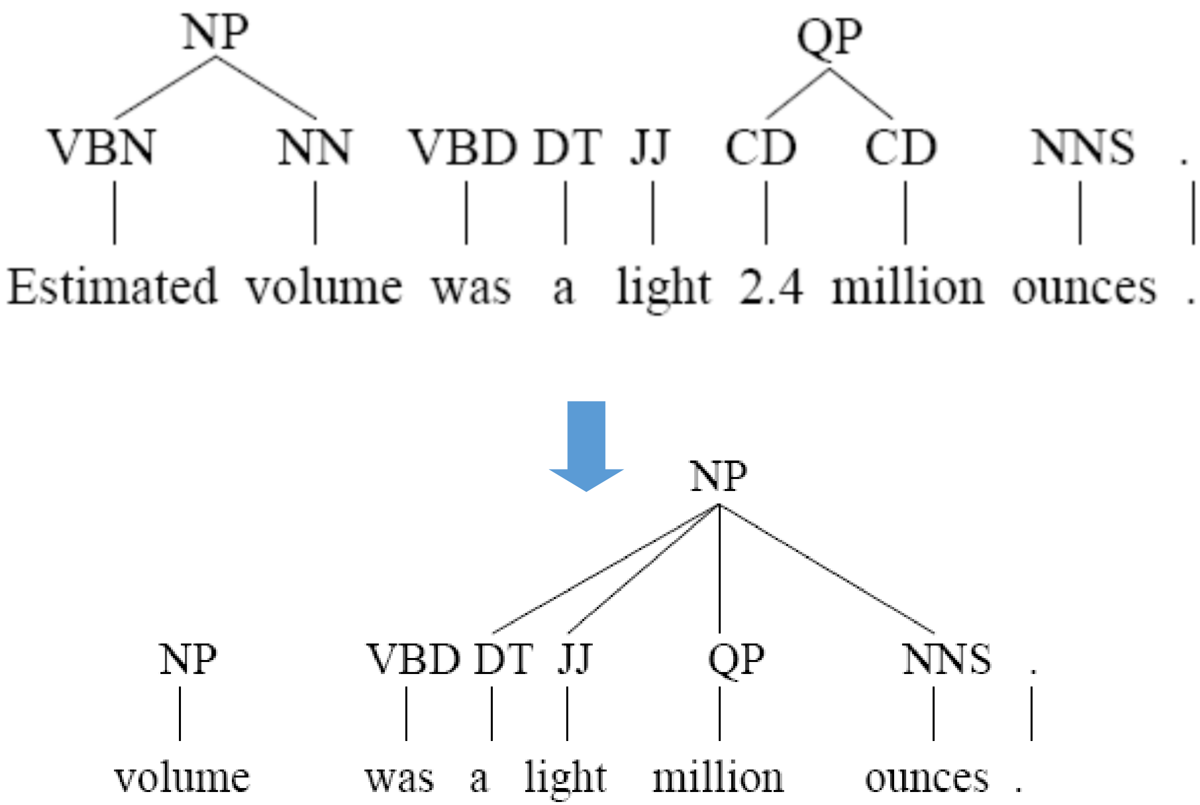
方法

- 看作序列标注问题

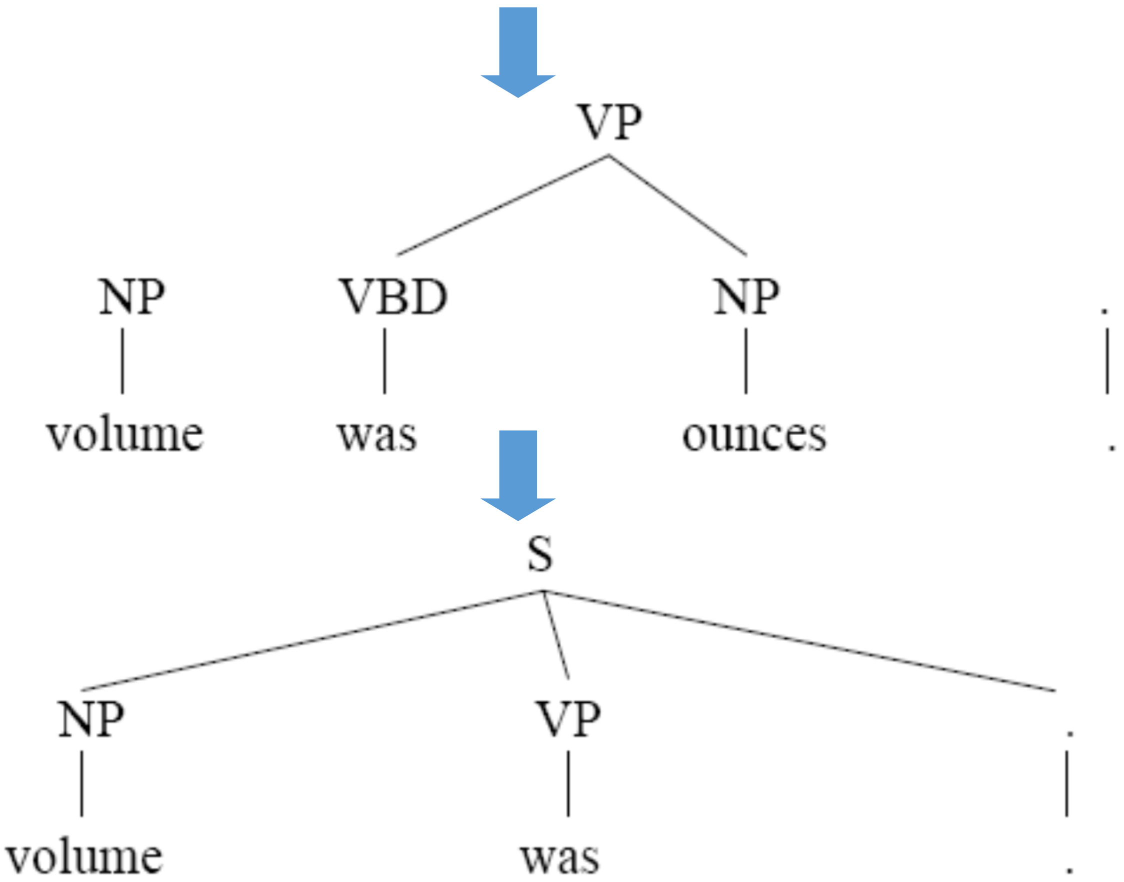
Integrated models	Cascade models 分类器序列
■ HMM	■ NB
■ MEMM	■ SVM
■ CRF	■ ME
NN models	■ Perceptron
■ LSTM+CRF	■ LR
■ BERT+CRF	

浅层句法分析

- 完全句法分析转化为chunking问题



浅层句法分析



浅层句法分析

- 从这个角度讲，序列结构预测问题可以转化为序列标注问题

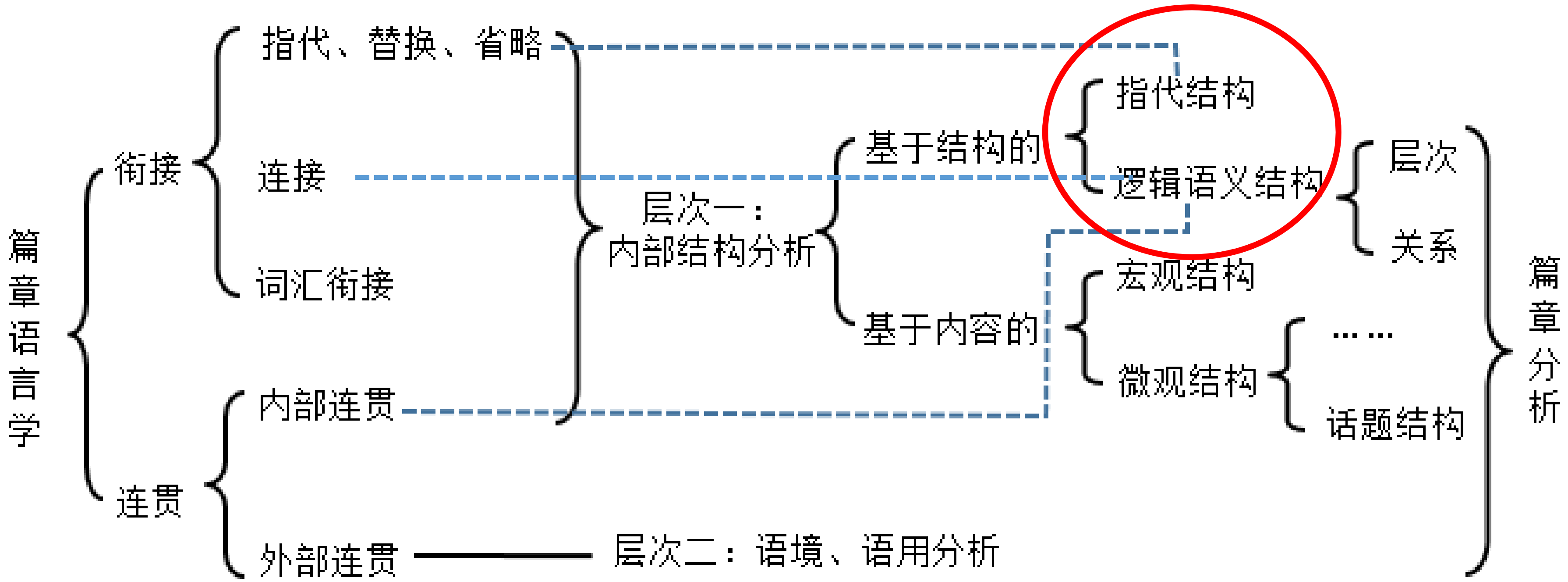
w	Estimated	volume	was	a	light	2.4	million	ounces	.
p	VBN	NN	VBD	DT	JJ	CD	CD	NNS	.
0	(NP	NP)				(QP	QP)		
0w		volume	was	a	light		million	ounces	.
0p		NP	VBD	DT	JJ		QP	NNS	.
1				(NP				NP)	
1w		volume	was					ounces	.
1p		NP	VBD					NP	.
2			(VP					VP)	
2w		volume	was						.
2p		NP	VP						.
3		(S							S)

大纲

- 句法分析
  - 短语结构句法分析
  - 依存结构句法分析
  - 浅层句法分析
- 篇章分析
  - 篇章结构分析

篇章结构分析

◻ 篇章分析


篇章语言学
衔接
指代、替换、省略
连接
词汇衔接
连贯
内部连贯
外部连贯
层次一：
内部结构分析
层次二：语境、语用分析
基于结构的
基于内容的
指代结构
逻辑语义结构
宏观结构
微观结构
层次
关系
……
话题结构
篇章分析

篇章结构分析

- 修辞结构理论(RST)与RST parsing
- PDTB与浅层篇章结构分析

修辞结构理论与RST parsing

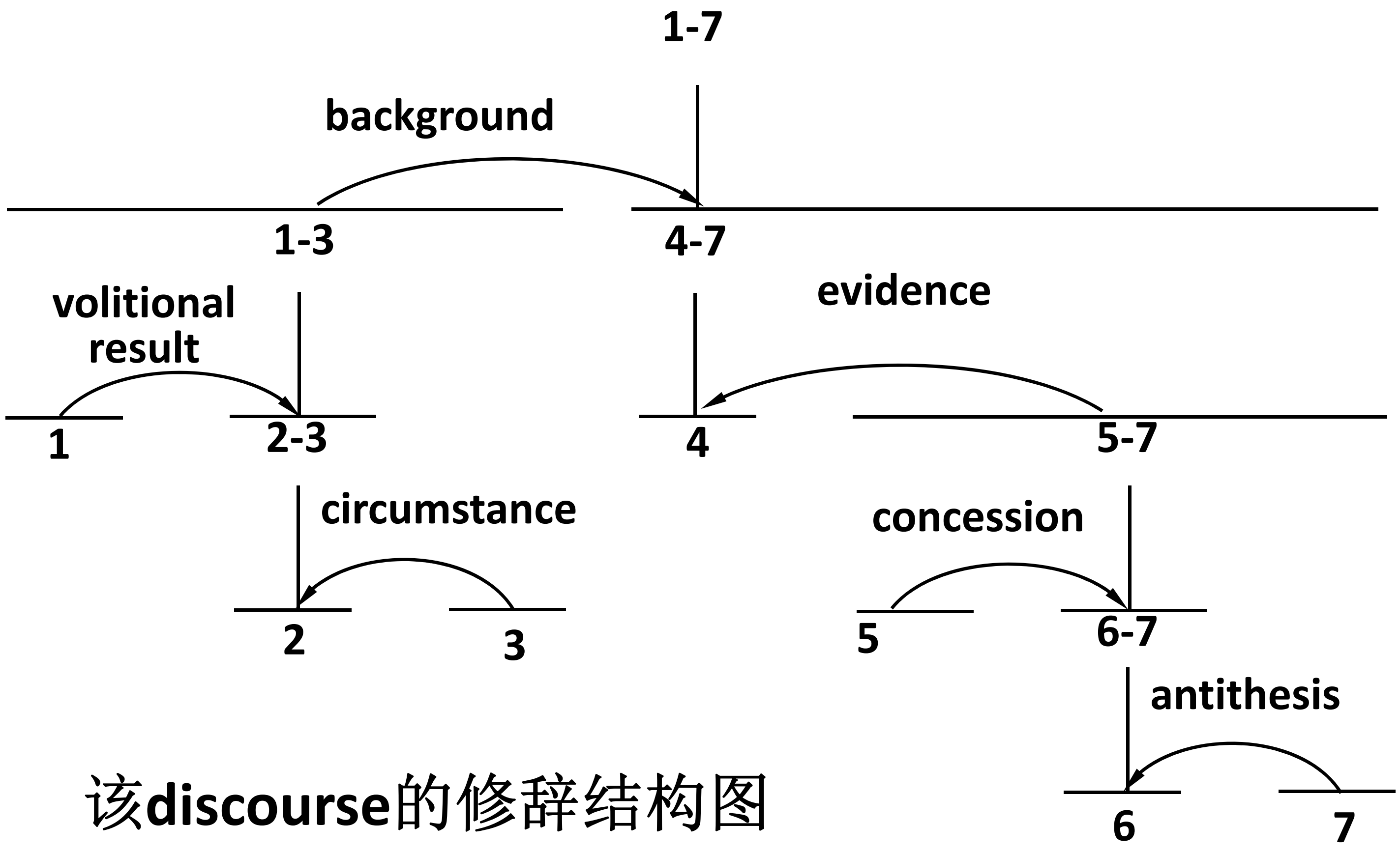
- Rhetorical structure theory (RST)
  - 定义了篇章基本单元(EDU)之间可能存在的23种修辞关系，存在修辞关系的两个篇章单元通常一个是核心(nucleus)一个是辅助(satellite)

修辞结构理论与RST parsing

1. Farmington police had to help control traffic recently
2. when hundreds of people lined up to be among the first applying for jobs at the yet-to-open Marriott Hotel.
3. The hotel's help-wanted announcement – for 300 openings – was a rare opportunity for many unemployed.
4. The people waiting in line carried a message, a refutation, of claims that the jobless could be employed if only they showed enough moxie.
5. Every rule has exceptions,
6. but the tragic and too-common tableaux of hundreds or even thousands of people snake-lining up for any task with a paycheck illustrates a lack of jobs,
7. not laziness.

划分为子句的一个**discourse**

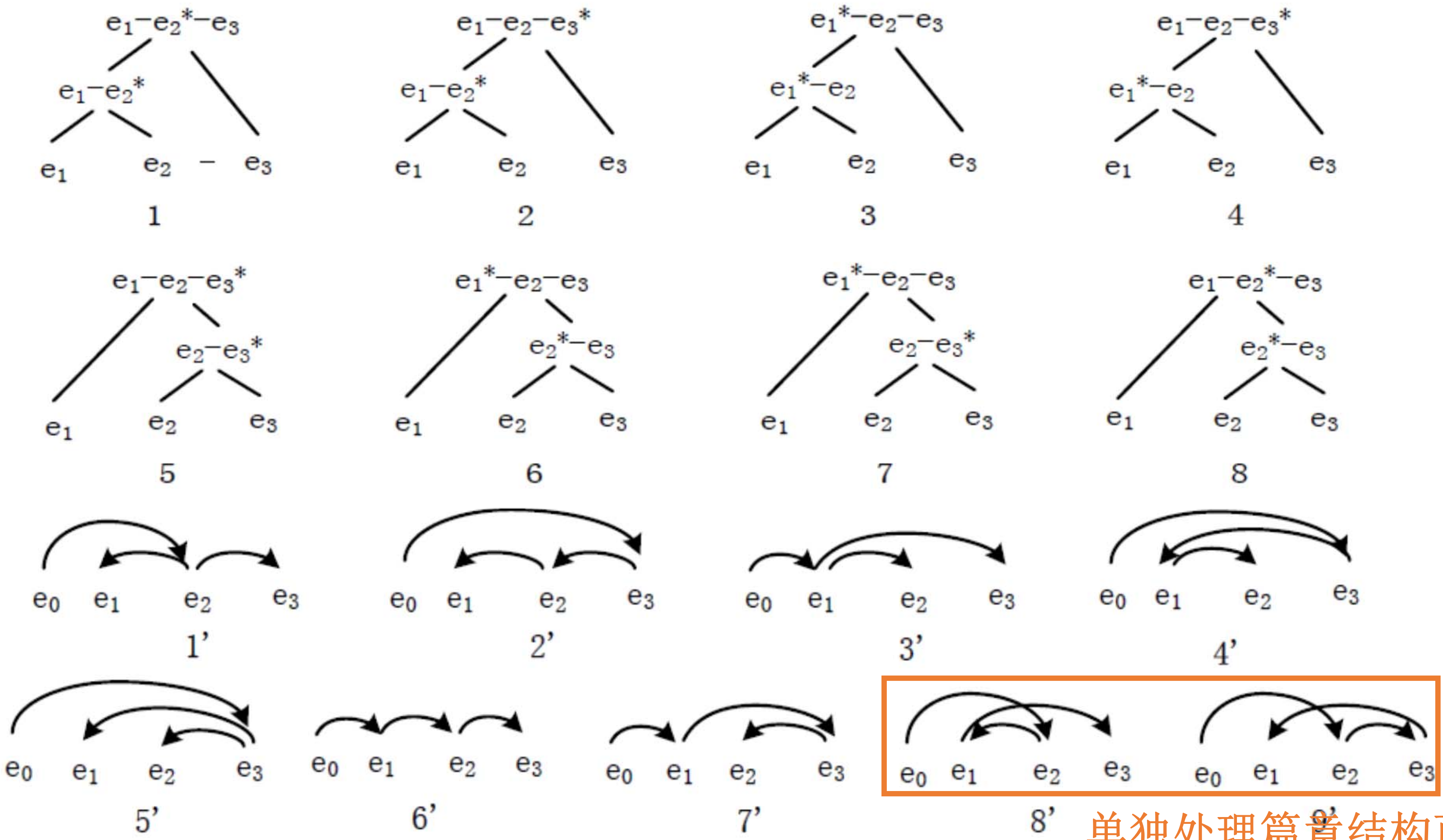
修辞结构理论与RST parsing



该discourse的修辞结构图

▫ 方法

- 可以看作句法分析的扩展



单独处理篇章结构可能出现，但不会出现在句法结构中的特殊情况

Using graph-based method to parse RST (Li et al., 2014)

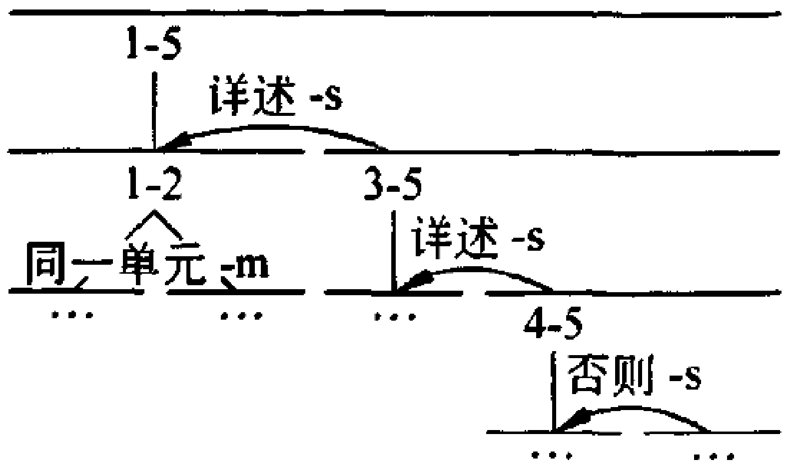
修辞结构理论与RST parsing

RST语料

- 英语：RST-DT， 2003 (基于PTB)

  http://www.isi.edu/~marcu/discourse/

  - **PTB**中的**385**篇文档
  - 以**clauses**或短语作为**EDU**
- 汉语
  - **CJPL,** 乐明**, 97**篇
  - **CDTB,** 苏大



篇章结构分析

- 修辞结构理论(RST)与RST parsing
- PDTB与浅层篇章结构分析

PDTB与浅层篇章结构分析

PDTB (Penn Discourse Treebank)

- 最popular的篇章树库
- 不标注discourse的整体结构，只是局部、浅层的两个单元的篇章关系
- 标注模式的基础是D-LTAG(篇章-词汇化树粘接文法)
- 将连接词看作二元篇章关系中的谓词，将篇章关系表示为由连接词连接的它的两个“论元”之间的关系

PDTB与浅层篇章结构分析

- PDTB中标注样本的例子
  - 显式关系**(explicit relation)**

Since **McDonald's menu prices rose this year,** *the actual decline may have been more***.**

  - 隐式关系**(implicit relation)**

$\boxed{\text{He said}}$ *more than 90 % of the funds were placed with Japanese institutional investors .* Implicit =AND **The rest went to investors from France and Hong Kong .**

Arg2 *Arg1* Connective $\boxed{\text{Attribution}}$

PDTB与浅层篇章结构分析

- PDTB中标注样本的例子
  - **AltLex (alternatively lexicalized)**

    And she further stunned her listeners by revealing her secret garden design method*: Commissioning a friend to spend "five or six thousand dollars . . . on books that I ultimately cutup."* AltLex **[After that], the layout had been easy.**
  - **EntRel (Entity based relation)**

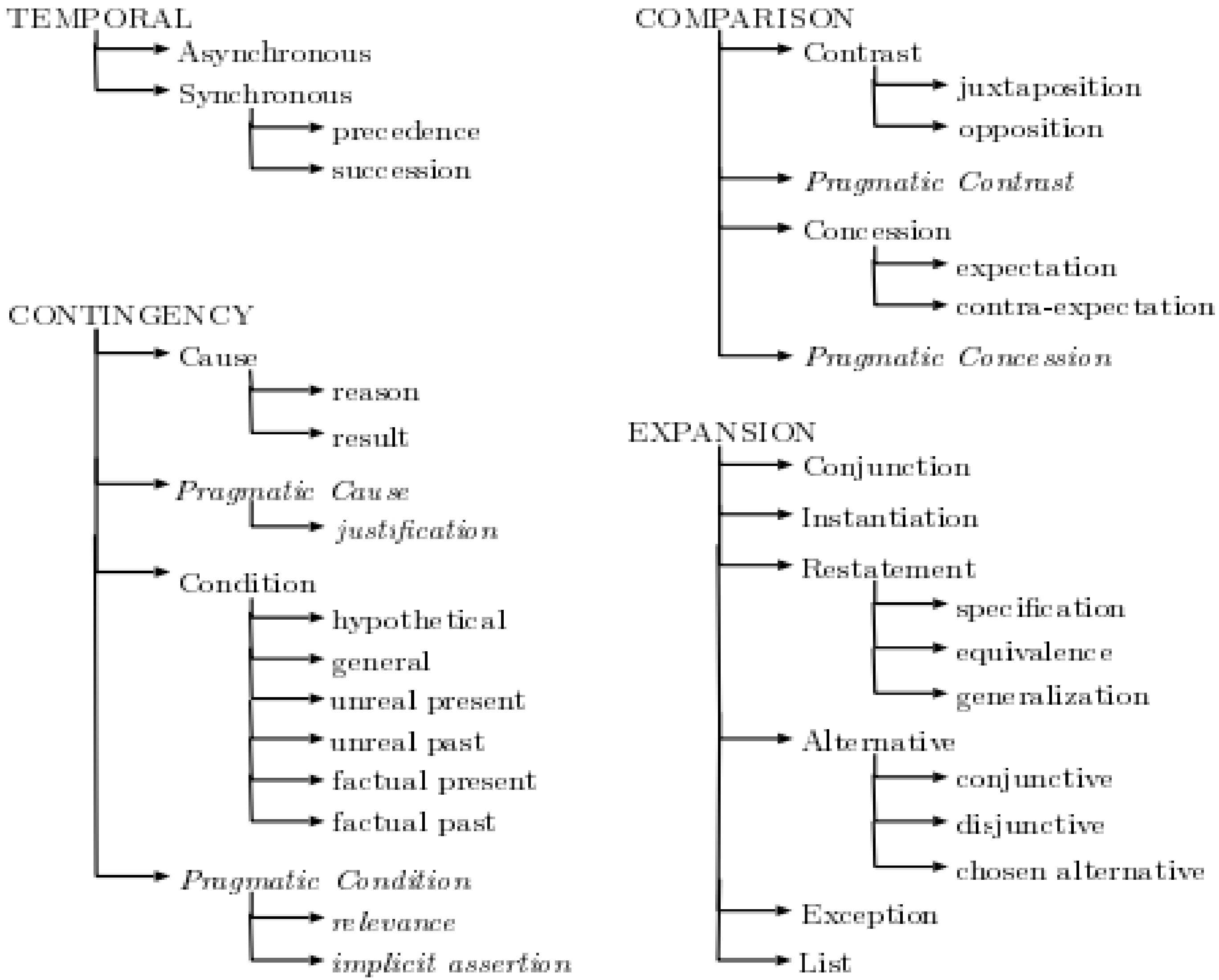
    Hale Milgrim, 41 years old, senior vice president, marketing at Elecktra Entertainment Inc., was named president of Capitol Records Inc., a unit of this entertainment concern. EntRel **Mr. Milgrim succeeds David Berman, who resigned last month.**

这两类占比非常少，有时忽略

PDTB与浅层篇章结构分析

- PDTB2.0 (2008)
  - 标注了PTB (WSJ)的2,312篇文档
  - 40,600个篇章关系
    - **18,459 explicit (45.5%)**
    - **16,053 implicit (39.5%)**

PDTB与浅层篇章结构分析



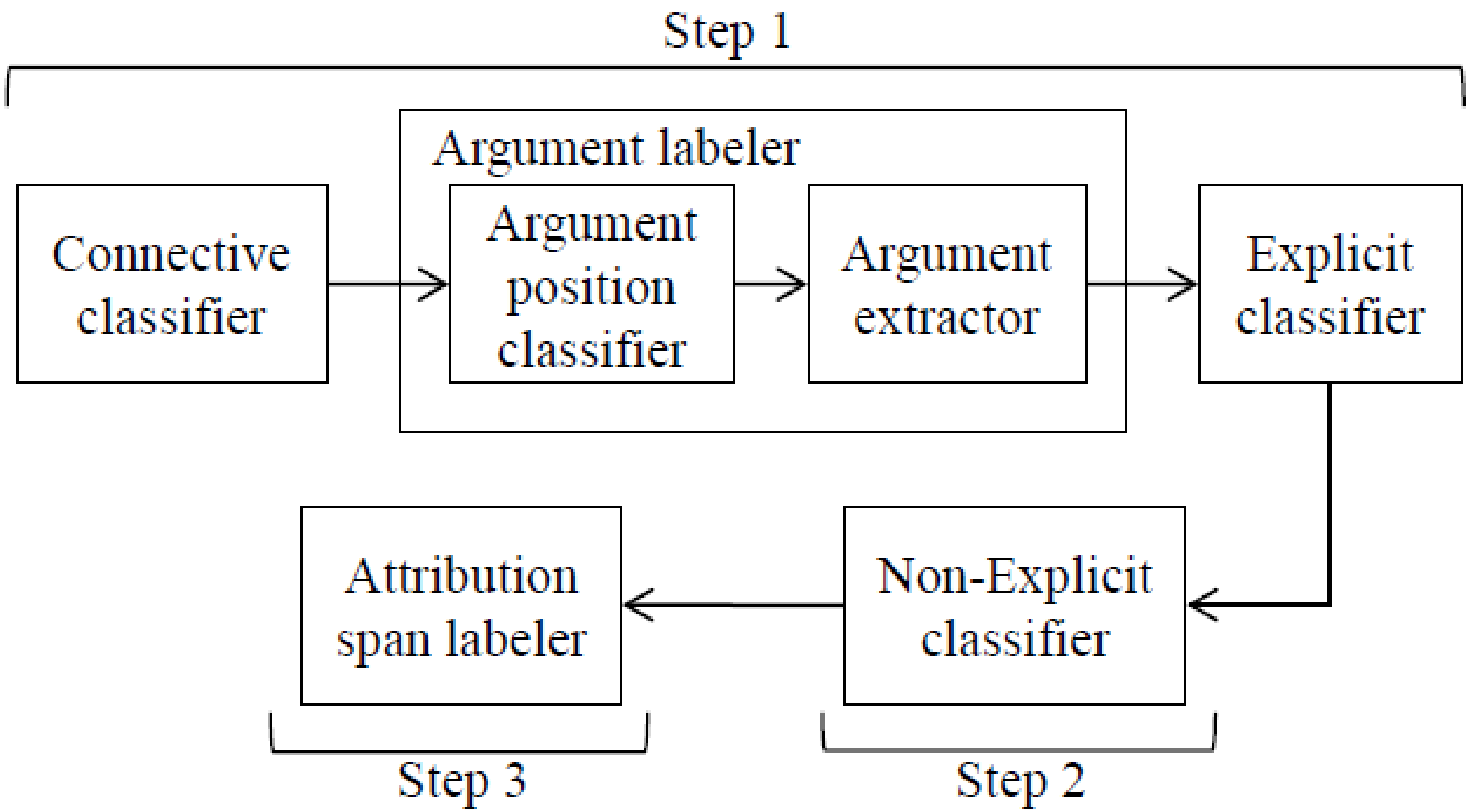
PDTB与浅层篇章结构分析

PDTB style汉语篇章标注语料

- CDTB: 在CTB上标注
  - **CDTB0.5 (2014)**
  - **CDTB1.0 (2016)**
    - **729 texts**
    - **11,023 discourse relations**
- HIT-CDTB: 在Ontonotes4.0上标注了525篇
  - **P3:**分句句间关系，句内关系；**72.01%**
  - **P2:**复句句间关系，句间关系；**26.28%**
  - **P1:**句群句间关系；**1.71%**

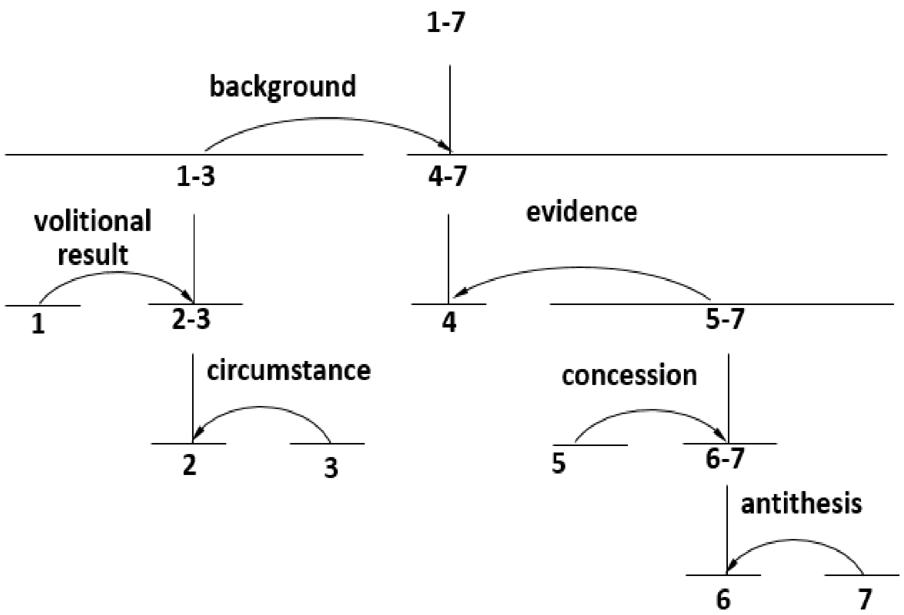
PDTB与浅层篇章结构分析

◻ PDTB上的结构分析任务



篇章结构分析

- 修辞结构理论(RST)与RST parsing
- PDTB与浅层篇章结构分析



Explicit

尽管房改的步伐在加快，
但福利分房的老办法仍未突破。

Implicit

平时要多晒晒太阳，
(因为)阳光中的紫外线能促进钙的吸收。

PDTB与浅层篇章结构分析

- 由于英语的表型性，其隐式关系绝大多数发生在两个句子之间，因此PDTB style parsing的主要困难：
  - 显式关系的论元划分
  - 隐式关系判别 ——看作一个典型的分类问题
    - 一个深度语义分析问题
    - 一个深度逻辑语义分析问题
    - 一个稀缺的深度逻辑语义分析问题

PDTB与浅层篇章结构分析

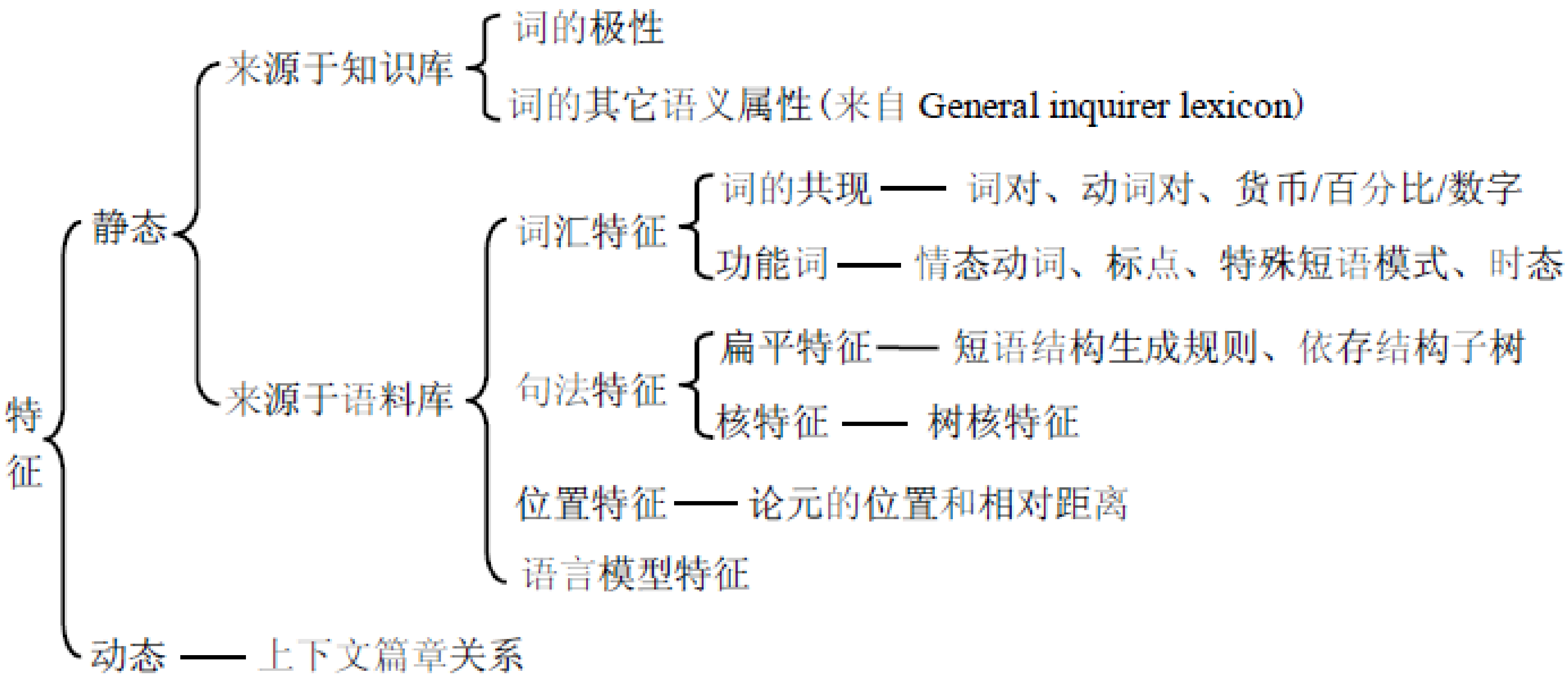
隐式篇章关系分析

- 早期特征工程方法
- 深度表示学习方法
- 论元间交互
- 显式关系的利用

PDTB与浅层篇章结构分析

早期特征工程方法

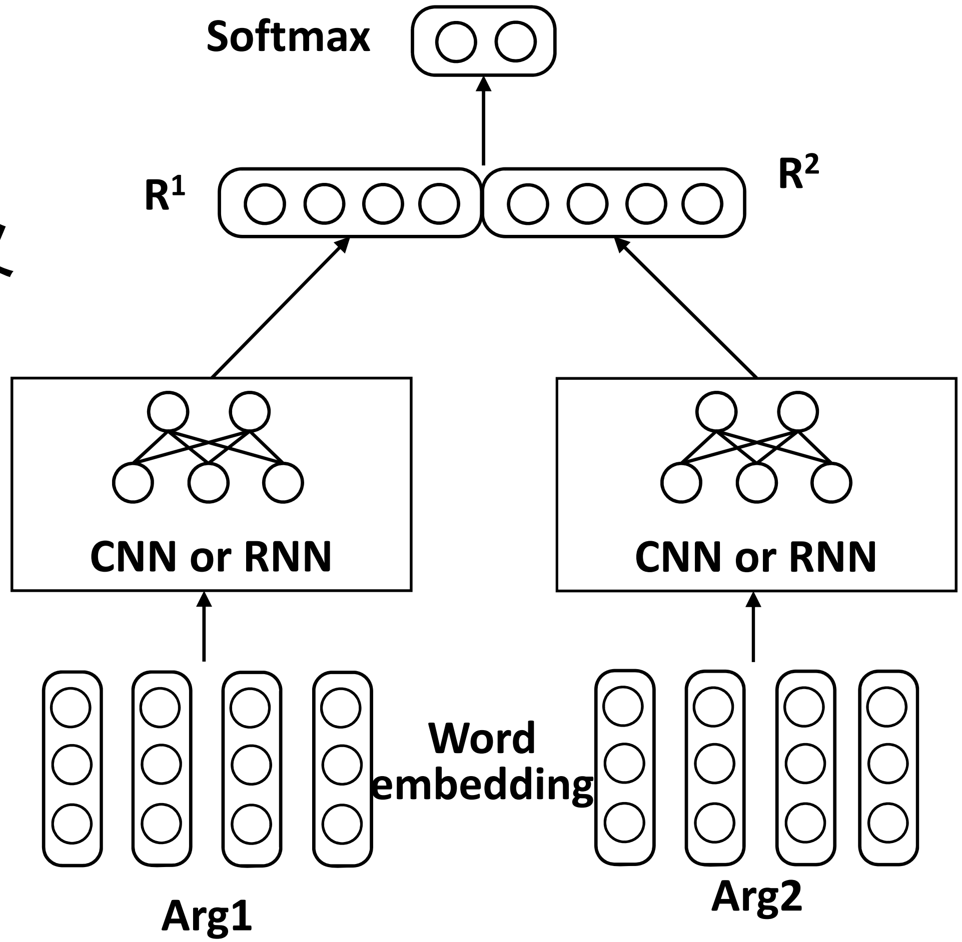
- 重点是有效特征的挖掘



PDTB与浅层篇章结构分析

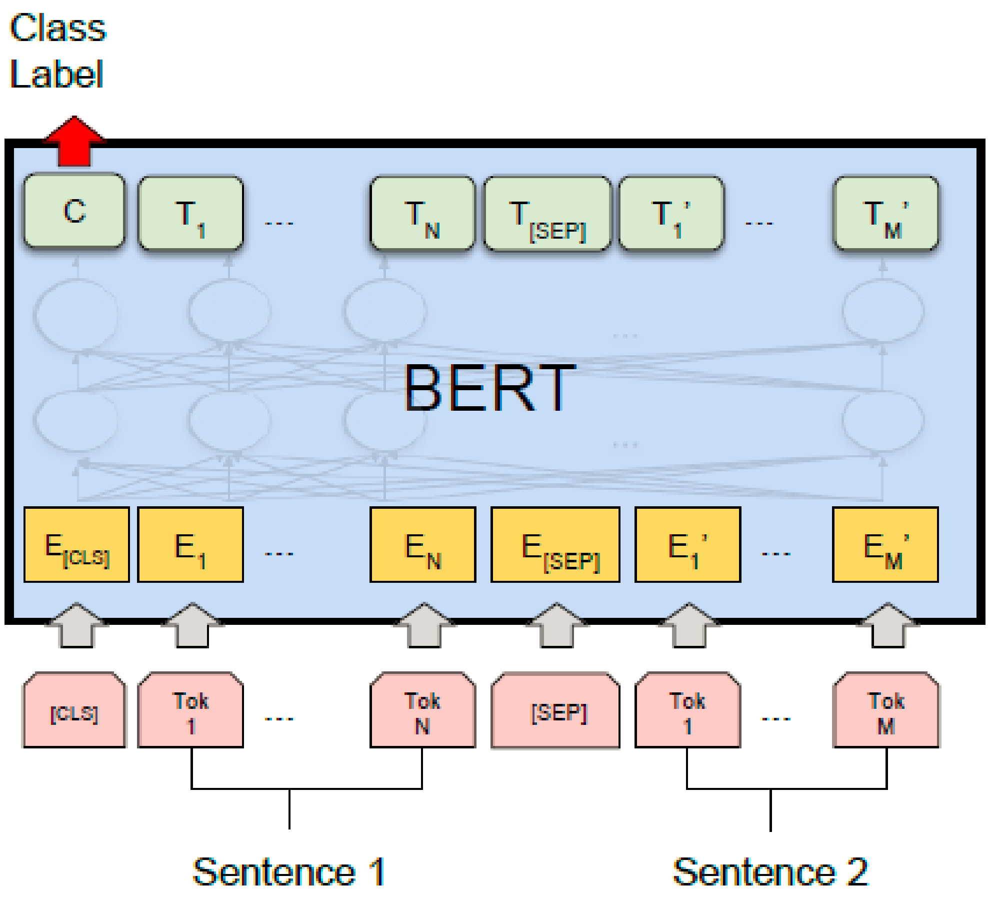
◘ 深度表示学习

- 希望词的表示、论元的表示包含逻辑语义
- 典型的双序列问题
- 也有工作直接学习与任务(篇章关系分析)相关的词汇向量表示



PDTB与浅层篇章结构分析

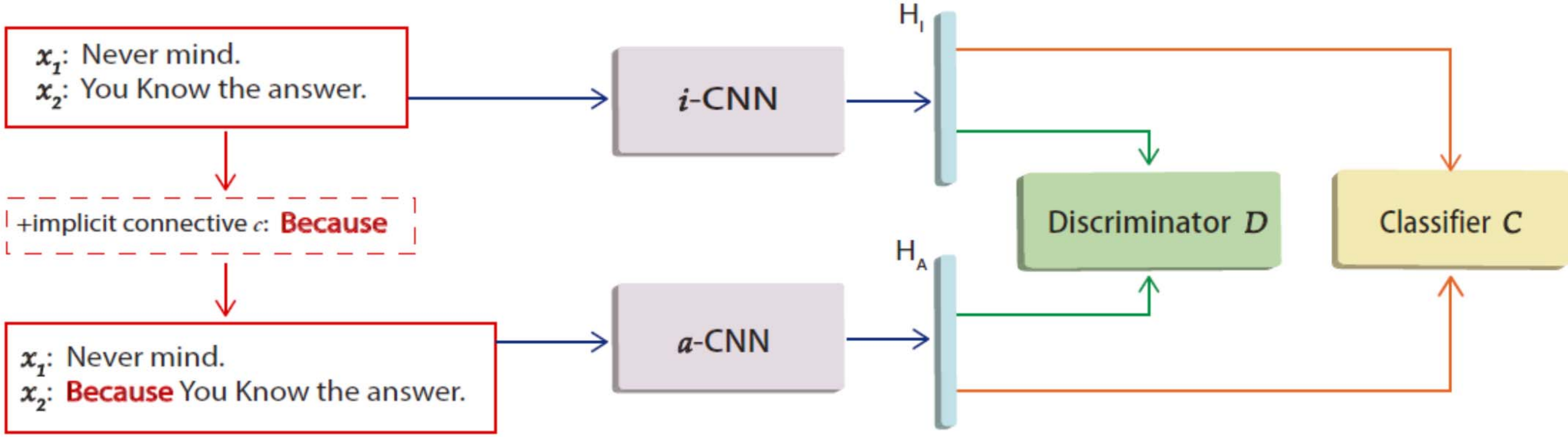
- 深度表示学习
  - 双序列分类是BERT的一个典型应用，可以直接finetune



PDTB与浅层篇章结构分析

深度表示学习

- 一个Adversarial模型
  - 没有连接词的隐式关系句对，其表示将越来越接近带有伪连接词的关系句对



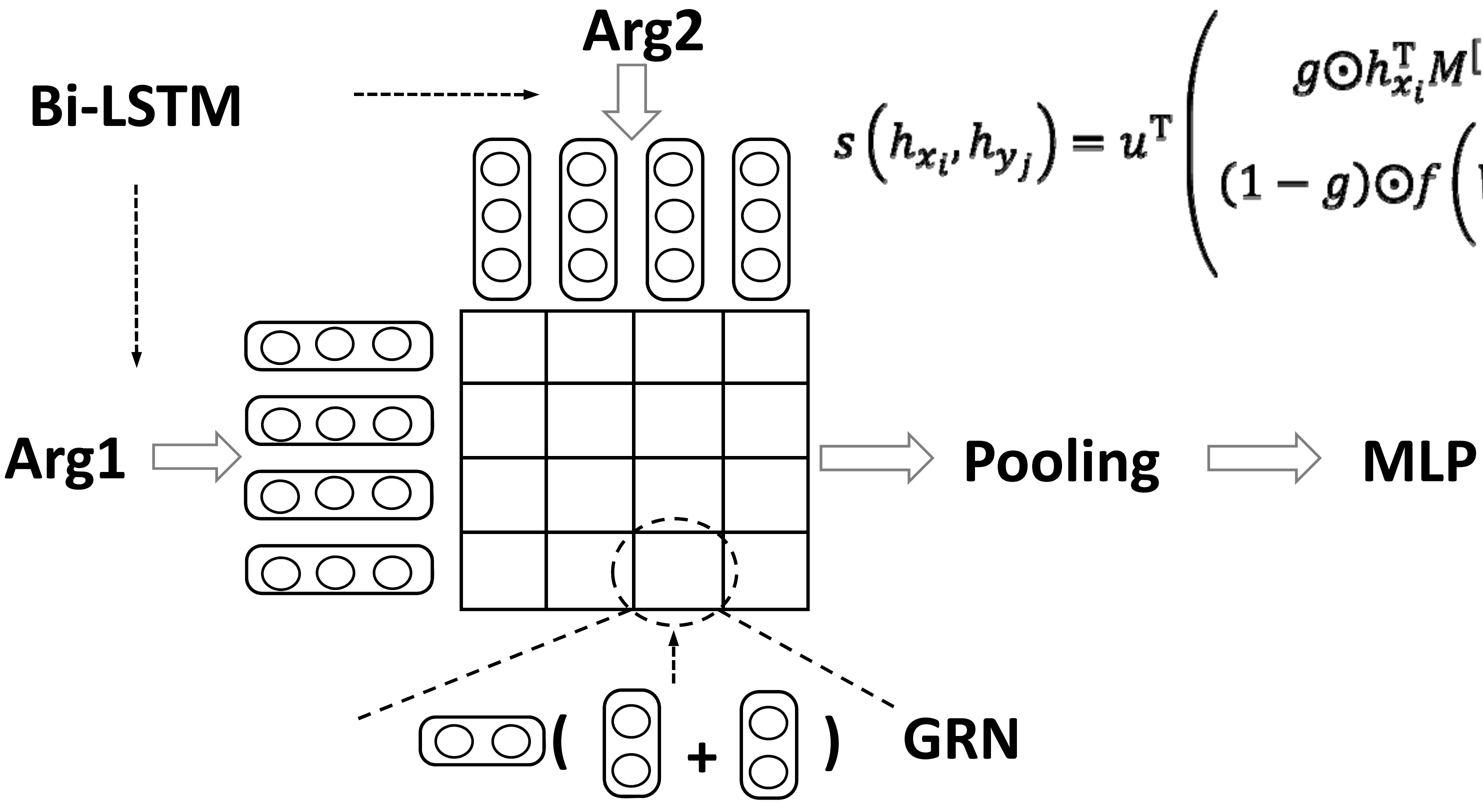
ACL2017: “Adversarial Connective-exploiting Networks for Implicit Discourse Relation Classification”

PDTB与浅层篇章结构分析

◘ 论元间交互 interaction

gated relevance network

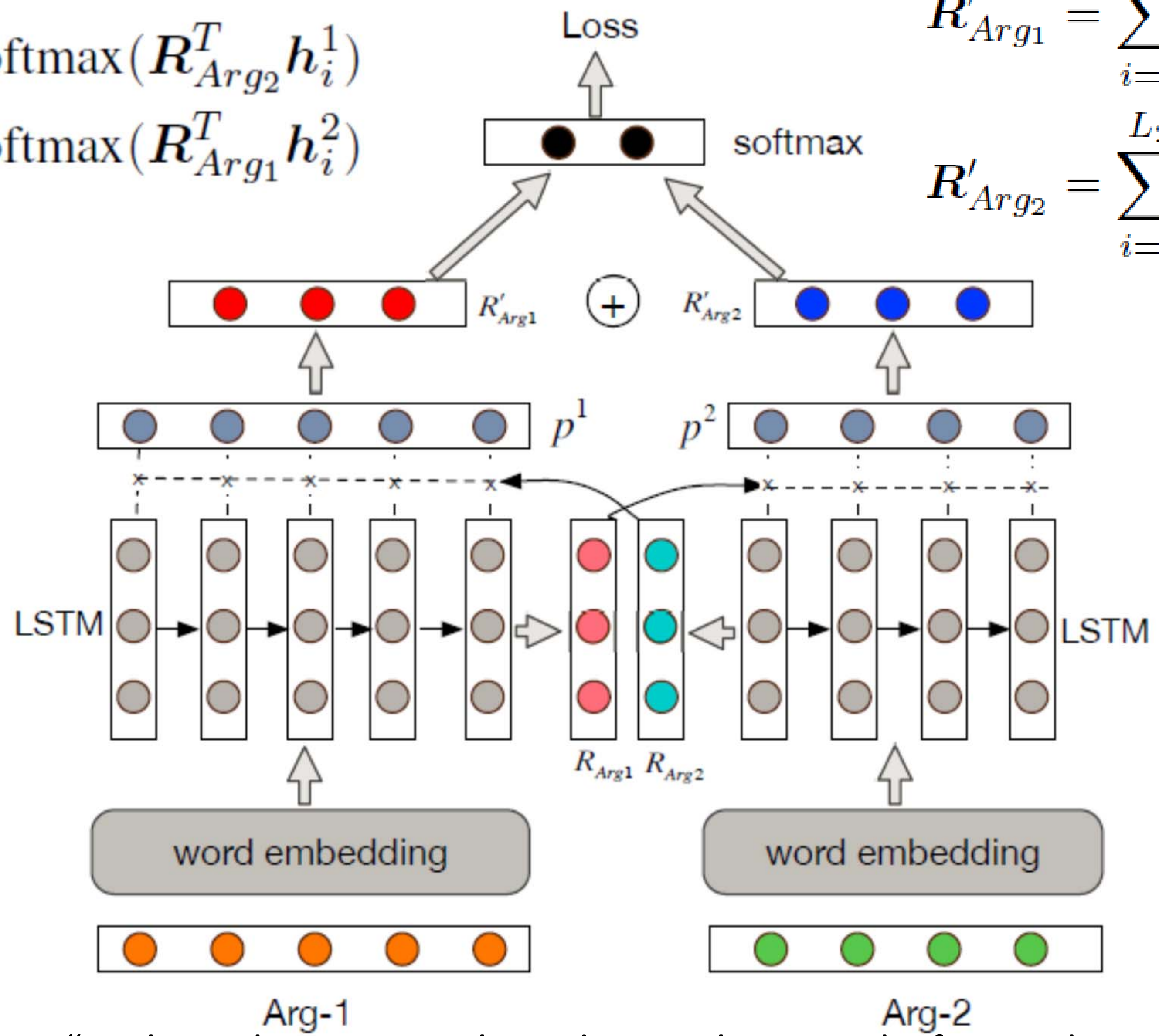
$$s\left(h_{x_i}, h_{y_j}\right) = u^{\mathrm{T}}\left(\begin{array}{c} g \odot h_{x_i}^{\mathrm{T}} M^{[1:r]} h_{y_j} + \\ (1-g) \odot f\left(V\begin{bmatrix} h_{x_i} \\ h_{y_j}\end{bmatrix}\right) + b \end{array}\right)$$



ACL2016: “Implicit discourse relation detection via a deep architecture with gated relevance network”

$$p_i^1 = \mathrm{Softmax}(\boldsymbol{R}_{Arg_2}^T \boldsymbol{h}_i^1)$$
$$p_i^2 = \mathrm{Softmax}(\boldsymbol{R}_{Arg_1}^T \boldsymbol{h}_i^2)$$

$$\boldsymbol{R}'_{Arg_1} = \sum_{i=0}^{L_1} \boldsymbol{h}_i^1 p_i^1$$
$$\boldsymbol{R}'_{Arg_2} = \sum_{i=0}^{L_2} \boldsymbol{h}_i^2 p_i^2$$

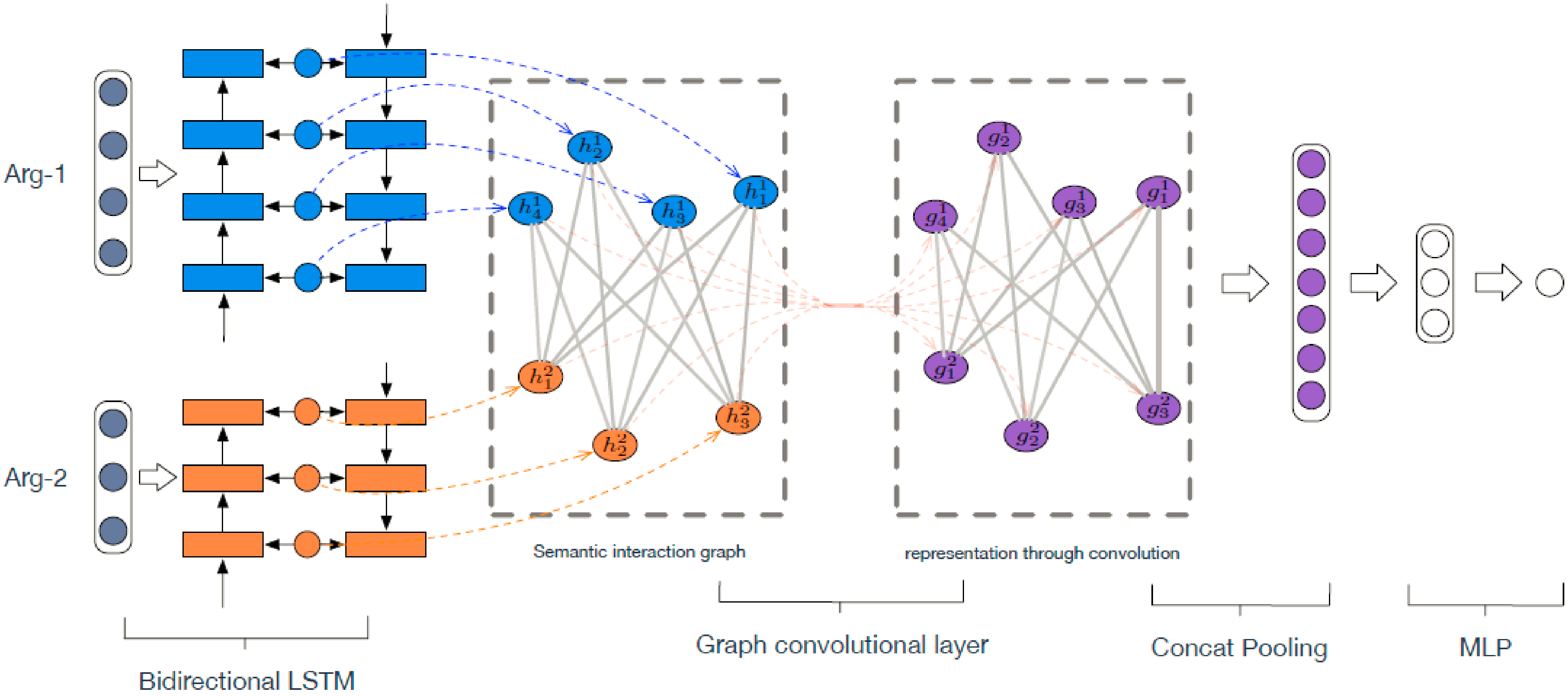


EMNLP2017: “Multi-task Attention-based Neural Networks for Implicit Discourse Relationship Representation and Identification”

PDTB与浅层篇章结构分析

◻ 论元间交互

- 基于GCN的交互



PDTB与浅层篇章结构分析

显式关系的利用

- 显式关系有连接词指示；
- 显式关系的判别准确率>90%；
- PDTB中的隐式关系，由标注者标注了伪显式连接词

✓ 思考：如何用显式关系或显式关系判别来辅助隐式关系判别？

PDTB与浅层篇章结构分析

- 隐式篇章关系分析关注模型对逻辑语义的建模
- 目前：
  - 预训练模型对逻辑语义的建模
    - 建模
    - 探测(**probe**)
  - 对文本coherence的建模
    - 评估
    - 生成

可参考**Lec2** 知识体系、问题及方法论**(**一**)**

本章对各知识点的要求

☐ 短语结构句法分析

- 掌握CYK原理和算法
- 以PCFG为例，掌握将规则模型加概率转化为概率统计模型的原理
- 结合n-gram, Naïve Bayes, HMM和PCFG，进一步加深对NLP中生成式模型的理解
- 理解RvNN和TreeLSTM解决短语结构预测的原理

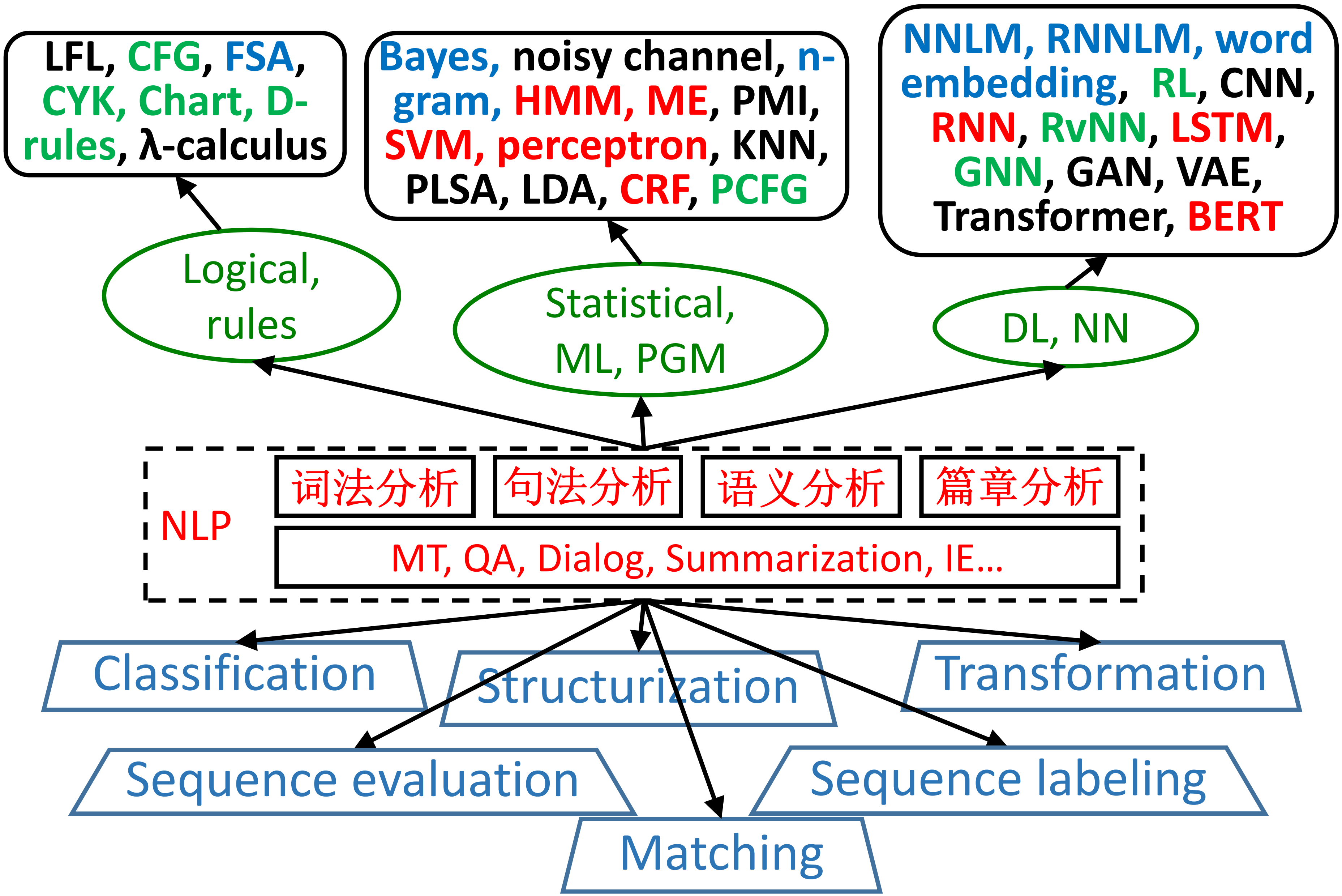
本章对各知识点的要求

依存结构分析

- 掌握用MST为依存结构建模的原理
- 通过MST所使用的线性判别函数，进一步加深对NLP中判别式方法的理解
- 以依存分析为例，掌握如何将结构预测转化为动作序列预测并用分类来解决的原理；并会使用transtition-based方法中的一种来转化依存结构

本章对各知识点的要求

- 浅层句法分析
  - 了解概念和原理
- 篇章结构分析
  - 了解两种篇章结构分析模式：RST和PDTB style
  - 掌握bi-sequence分类的常规做法



Tasks, problems, methodologies and models in NLP

Reference

- M. Collins. Three generative, lexicalized models for statistical parsing. In Proceedings of ACL-EACL1997.
- R. Socher et al. Parsing natural scenes and natural language with recursive neural networks. In Proc. of ICML2011.
- K. Tai et al. Improved semantic representations from tree-structured long short-term memory networks. In Proc. of ACL2015.
- D. Yogatama et al. Learning to compose words into sentences with reinforcement learning. In Proc. of ICLR2017.
- J. Choi et al. Learning to compose task-specific tree structures. In Proc. of AAAI2018.
- A. Vaswani et al. Attention is all you need. In Proc. of NIPS2017.
- Yoav Goldberg. 2019. Assessing BERT's syntactic abilities. CoRR, abs/1901.05287.
- What does BERT learn about the structure of language. In Proc. of ACL2019.
- X. Li and J. Eisner. Specializing word embeddings (for parsing) by information bottleneck. In Proc. of EMNLP2019.

Reference

- J. Eisner. Three new probabilistic models for dependency parsing: An exploration. In Proceedings of COLING1996.
- R. McDonald and F. Pereira. Online learning of approximate dependency parsing algorithms. In Proceedings of EACL2006.
- R. McDonald et al. Online large-margin training of dependency parsers. In Proceeding of ACL2005.
- J. Nivre. Incrementality in deterministic dependency parsing. In Proceedings of ACL2004.
- H. Yamada and Y. Matsumoto. Statistical dependency analysis with support vector machines. In Proceedings of IWPT2003.
- W. Wang and B. Chang. Graph-based dependency parsing with bidirectional LSTM. In Proc. of ACL2016.
- T. Dozat and C. Manning. Deep Biaffine attention for neural dependency parsing. In Proc. of ICLR2017.
- T. Ji et al. Graph-based dependency parsing with graph neural networks. In Proc. of ACL2019.
- D. Marcheggiani and I. Titov. Encoding sentences with graph convolutional networks for semantic role labeling. In Proc. of EMNLP2017.

Reference

- **D. Chen and C. Manning. A fast and accurate dependency parser using neural networks. In Proc. of EMNLP2014.**
- **C. Dyer et al. Transition-based dependency parsing with stack long short-term memory. In Proc. of ACL2015**
- **E. Kiperwasser and Y. Goldberg. Simple and accurate dependency parsing using bidirectional LSTM feature representations. TACL2016.**
- **G. Lample. Neural architectures for named entity recognition. In Proc. of NAACL2016.**
- **B. Webber and A. Joshi. 1998. Anchoring a lexicalized tree-adjoining grammar for discourse. In Coling-ACL Workshop on Discourse Relations and Discourse Markers, pp: 86–92.**
- 乐明**. 2008.** 汉语篇章修辞结构的标注研究**.** 中文信息学报**, 22(7): 19-23.**
- **R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. 2008. The Penn Discourse TreeBank 2.0. In Proceedings of LREC.**
- **Z. Lin, H. Ng, and M.-Y. Kan. 2011. A PDTB-styled end-to-end discourse parser. Technical Report TRB8/10. School of Computing, National University of Singapore.**

Reference

- **Z. Lin, M.-Y. Kan, and H. Ng. 2009. Recognizing implicit discourse relations in the Penn Discourse Treebank. In Proceedings of EMNLP.**
- **E. Pitler and A. Nenkova. 2009. Using syntax to disambiguate explicit discourse connectives in text. In Proceedings of the ACL-IJCNLP Conference Short Papers.**
- **J. Chen, Q. Zhang, P. Liu, X. Qiu and X. Huang. 2016. Implicit discourse relation detection via a deep architecture with gated relevance network. In Proceedings of ACL.**
- **M. Lan, J. Wang, Y. Wu, Z. Niu, H. Wang. 2017. Multi-task Attention-based Neural Networks for Implicit Discourse Relationship Representation and Identification. In Proceedings of EMNLP, pages 1310-1319.**
- **L. Qin, Z. Zhang, H Zhao, Z. Hu and E. P. Xing. 2017. Adversarial Connective-exploiting Networks for Implicit Discourse Relation Classification. In Proceedings of ACL, pages 1006-1017.**
- **A. Laha and V. Raykar. 2016. An empirical evaluation of various deep learning architectures for bi-sequence classification tasks. In Proc. of Coling.**

Reference

- J. Zhang, H. Luan, M. Sun, F. Zhai, J. Xu, M. Zhang and Y. Liu. Improving the transformer translation model with document-level context. In Proceedings of EMNLP, 2018: 533-542.
- E. Voita, P Serdyukov, R. Sennrich and I Titov. Contex-aware neural machine translation learns anaphora resolution. In Proceedings of ACL, 2018: 1264-1274.
- S. Maruf and G. Haffari. Document context neural machine translation with memory networks. In Proceedings of ACL, 2018: 1275-1284.