

2.7 语法解析（下）

林洲汉
上海交大电院

2024年秋季学期

- ▶ **依存式语法 (dependency grammar) 简介**
 - ▶ 构成式语法的问题
 - ▶ 依存式语法基本概念
 - ▶ 依存式语法与构成式语法的互转条件
- ▶ **使用shift-reduce parser做依存式语法解析**
- ▶ **评价指标**
- ▶ **常用工具**

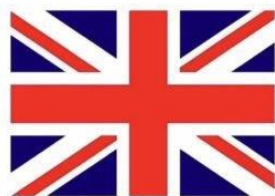
- ▶ **依存式语法 (dependency grammar) 简介**
 - ▶ 构成式语法的问题
 - ▶ 依存式语法基本概念
 - ▶ 依存式语法与构成式语法的互转条件
- ▶ 使用shift-reduce parser做依存式语法解析
- ▶ 评价指标
- ▶ 常用工具

构成式语法的问题



分析语

做



do, doing
did, done

向分析语转变中的屈折语

делать, делаю, делаешь, делает,
делаем, делают, делают, делал
делала, делала, делали, делайте, дел
ающий, делающе, делающему
делающим, делающему, делающая,
делающей, делающую, делающей,
делающее,
делающие, делающие, делающими,
делавший, делавшего, делавшему,
делавшим, делавшем, делавшая,
делавшей, делавшую,
делавшей, делавшее, делавшие,
делавших, делавшими, делаться,
делаюсь, делаешься, делается,
делается, делаетесь, делаются,
делался, делалось, делалось,
делались, делались, делайтесь,
делаемый, делаемого, делаемому, д
елаемы, делаемом,
делаем, делаемая, делаемой, делае
мую, делаемую, делаем, делаемое,
делаемо, делаемые, делаемым
делаемыми, делаемы,
делающийся, делающегося, делаю
щемся, делающимся, делающемся,
делающаяся, делающейся, делающе
юся, делающеюся, делающеся, дел
ающиеся, делающихся, делающими
ся, деланный, деланного,
деланному, деланным,
деланном, делан, деланная, деланн
ой, деланную, деланною, делана, дел

硬核屈折语

构成式语法的问题

分析语



他是个好人。 ✓
是个好人他。 ✓
好人他是个。 ✓
好人是个他。 ✗
是他个好人。 ✗
是好人个他。 ✗
..... ✗

向分析语转变中的屈折语



He's a good person. ✓
A good person he is. ✓
A person he good is. ✗
Person a he good is. ✗
Good person a he is. ✗
Good he a person is. ✗
..... ✗

硬核屈折语



Он хороший человек. ✓
Он человек хороший. ✓
Хороший человек он. ✓
Хороший он человек. ✓
Человек хороший он. ✓
Человек он хороший. ✓

Он	他
Хороший	好
Человек	人

所有 A_3^3 种全排列，全是对的！

构成式语法的问题

硬核屈折语

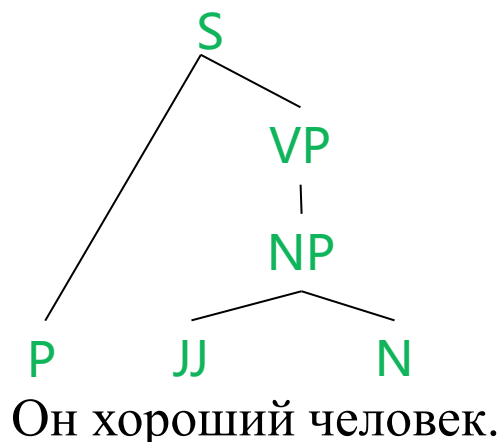


Он
Хороший
Человек

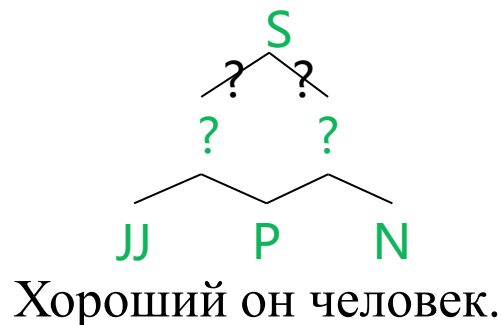
他
好人



Constituent的种类随着句子语序的变化指数型增长.....



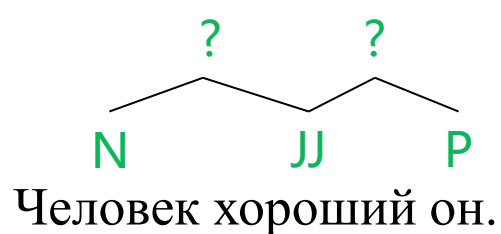
✓



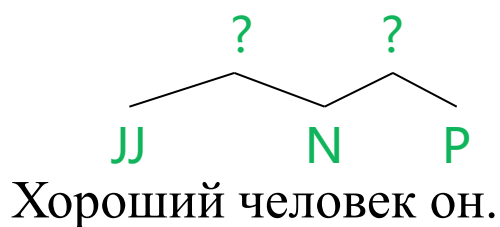
✓



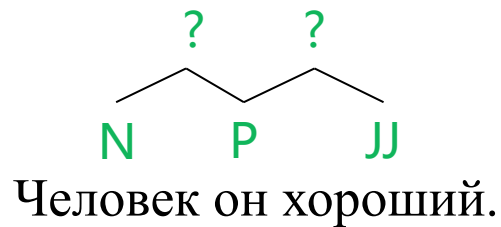
✓



✓



✓



✓

构成式语法的问题

硬核屈折语

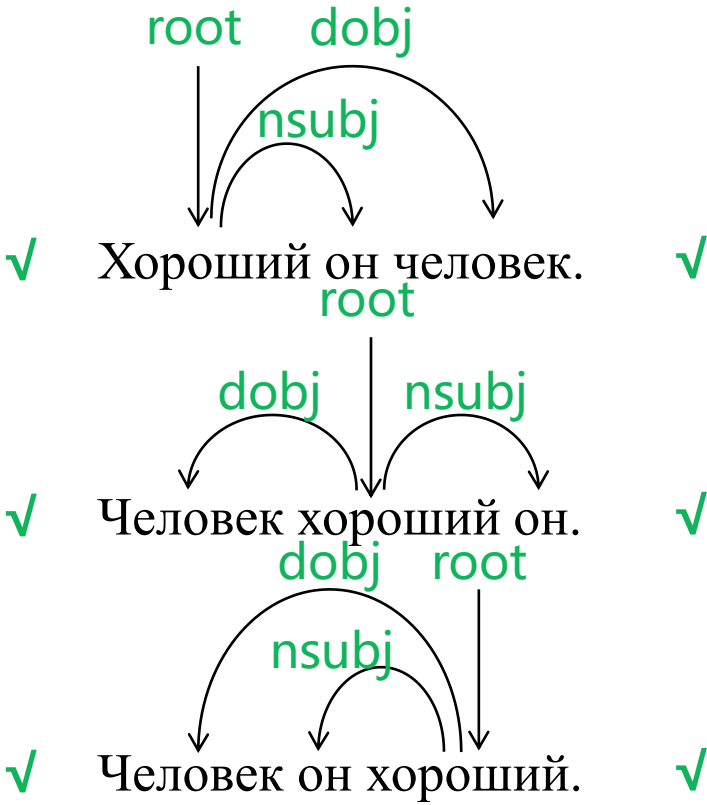
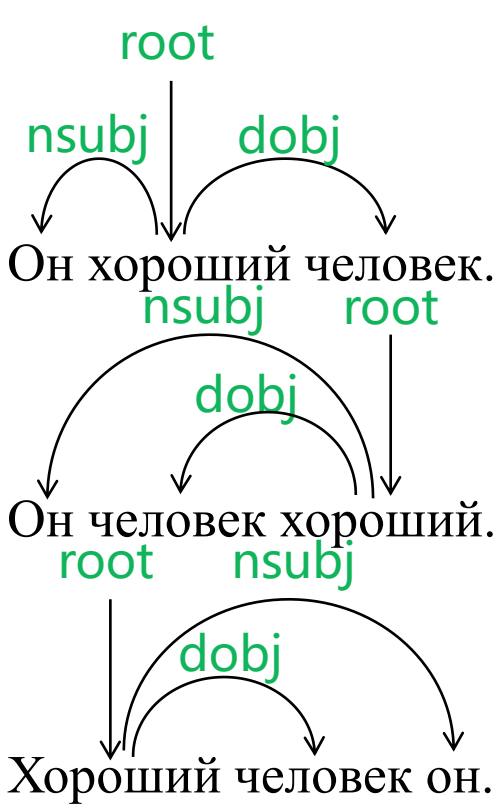


Он
Хороший
Человек

他
好
人

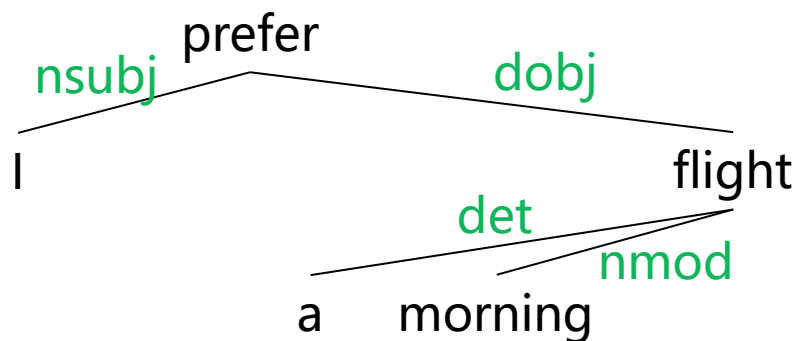
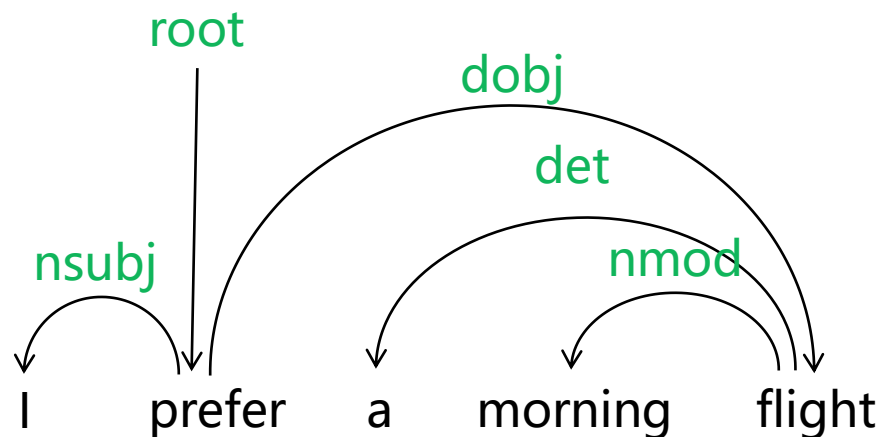
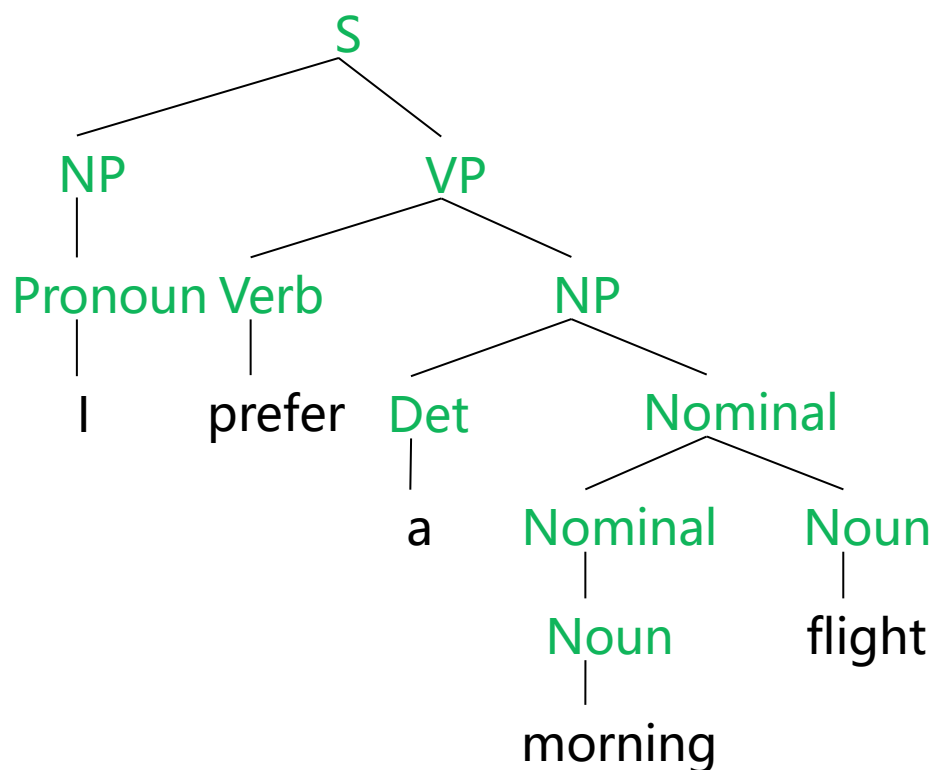


词之间关系的类型不随语序的改变而改变。

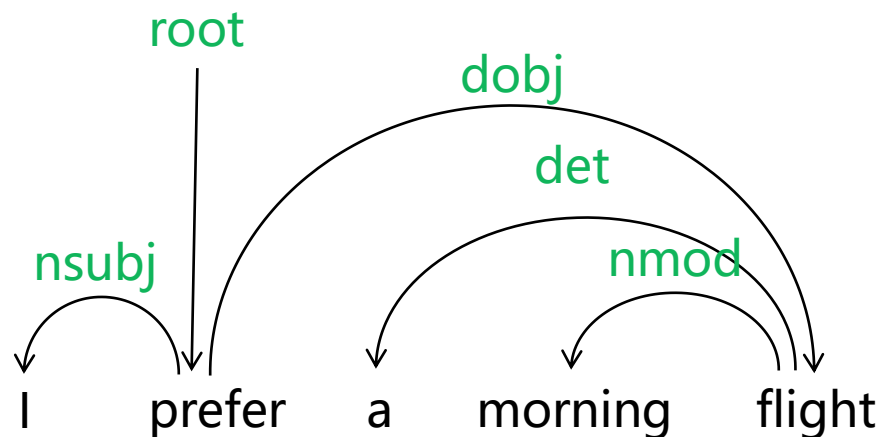


依存式语法 (dependency grammar) 简介

I prefer a morning flight.



依存式语法 (dependency grammar) 简介



依存关系 (dependency relation) 带标签的有向箭头，两边分别是head & dependent

NSUBJ(名词性主语)

CCOMP(补语从句)

NMOD(名词性修饰语)

NUMMOD(数词修饰语)

CASE(介词与格标记)

DOBJ(直接宾语)

XCOMP(开句补语)

AMOD(形容词性修饰语)

APPOS(同位修饰语)

CONJ(连词)

IOBJ(间接宾语)

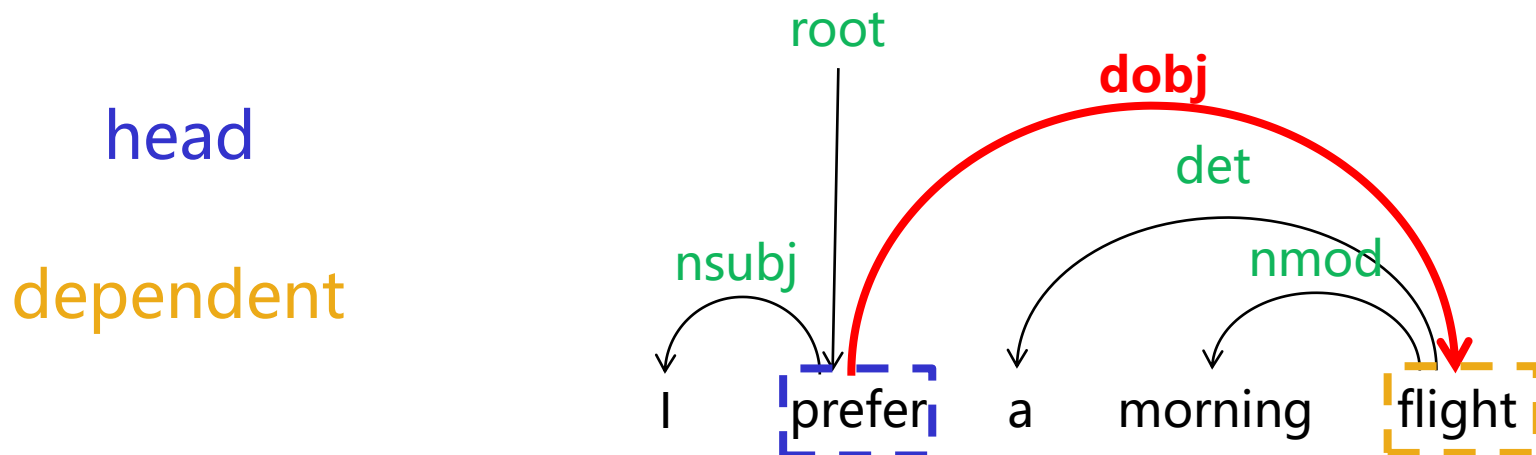
DET(冠词)

CC(并列连词)

根节点 (root)

词节点 (vertex)

依存式语法 (dependency grammar) 简介



依存关系 (dependency relation) 带标签的有向箭头，两边分别是head & dependent

NSUBJ(名词性主语)

CCOMP(补语从句)

NMOD(名词性修饰语)

NUMMOD(数词修饰语)

CASE(介词与格标记)

DOBJ(直接宾语)

XCOMP(开句补语)

AMOD(形容词性修饰语)

APPOS(同位修饰语)

CONJ(连词)

IOBJ(间接宾语)

DET(冠词)

CC(并列连词)

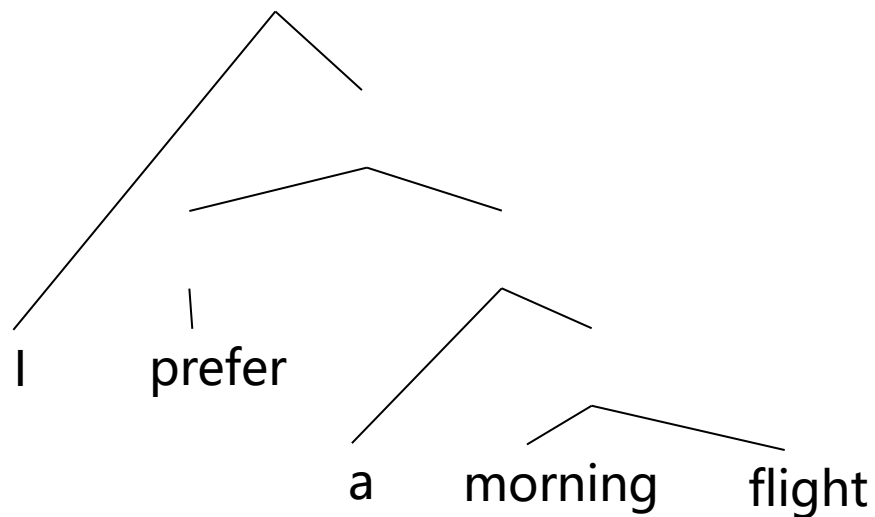
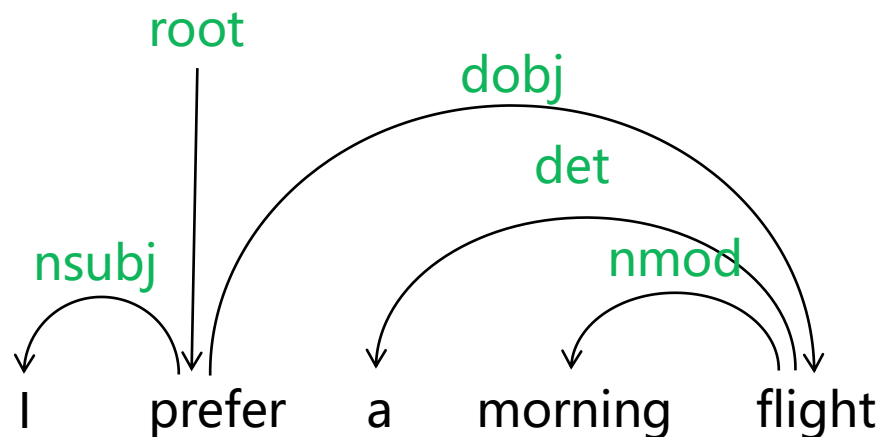
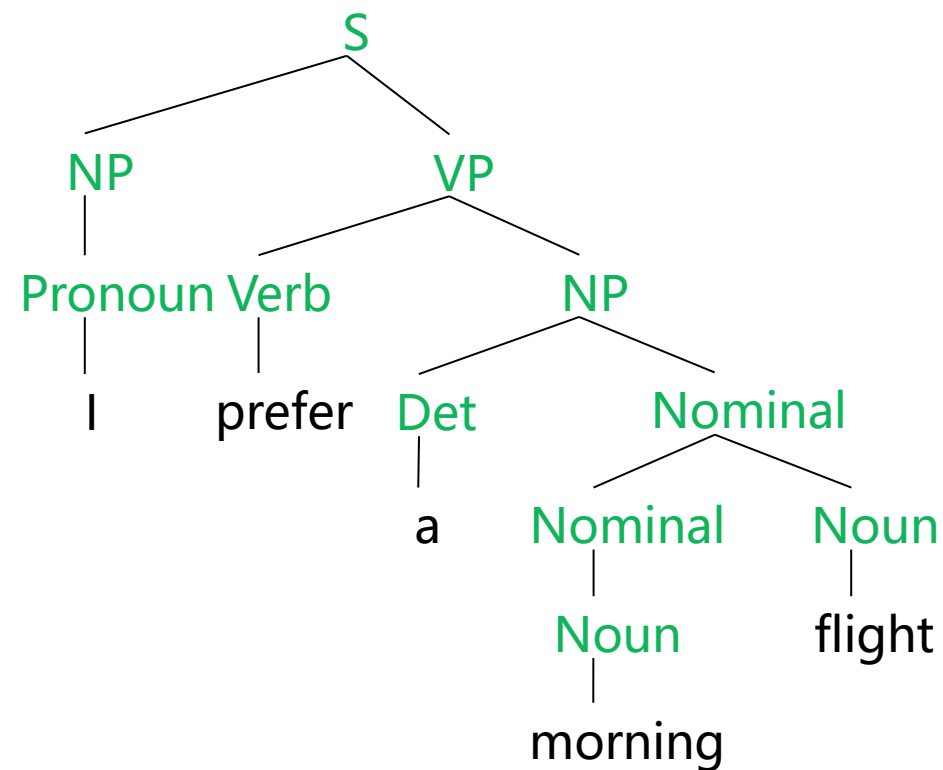
根节点 (root) 每个句子有且仅有一个根节点，且根节点只能作为head

词节点 (vertex)

1. 每个词节点只能有一条连入边
2. 从根节点到任意词节点的路径唯一

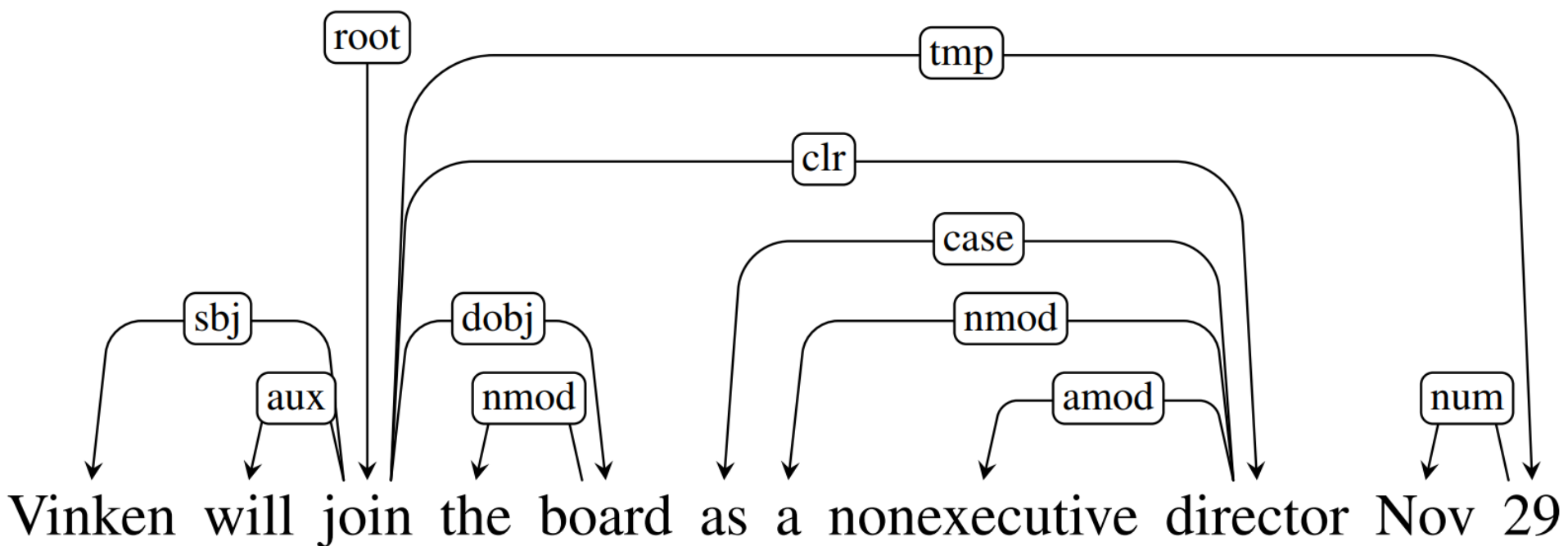
依存式语法与构成式语法的互转

I prefer a morning flight.



依存式语法与构成式语法的互转

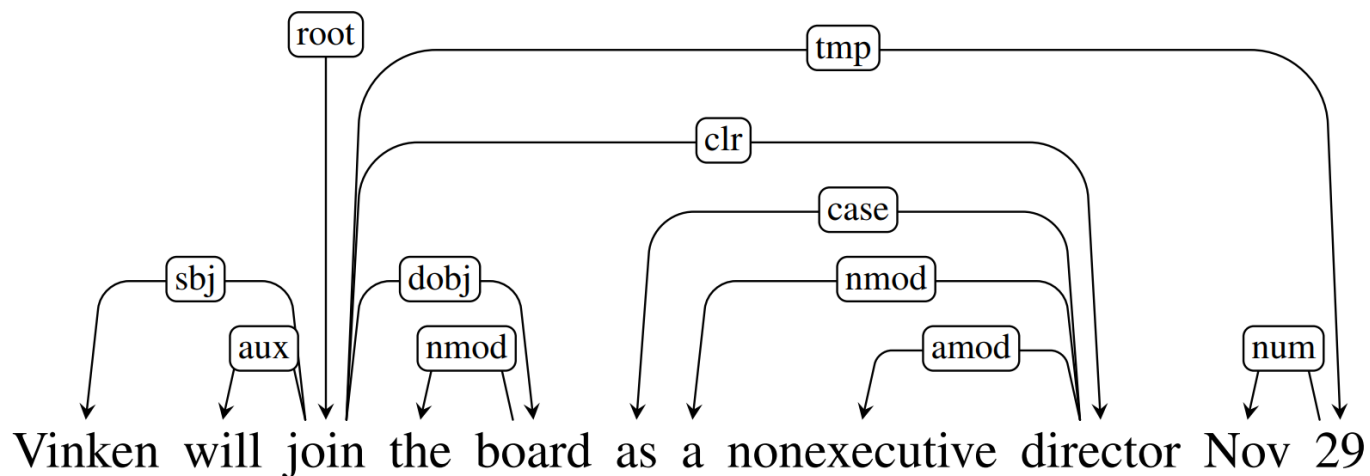
例子:



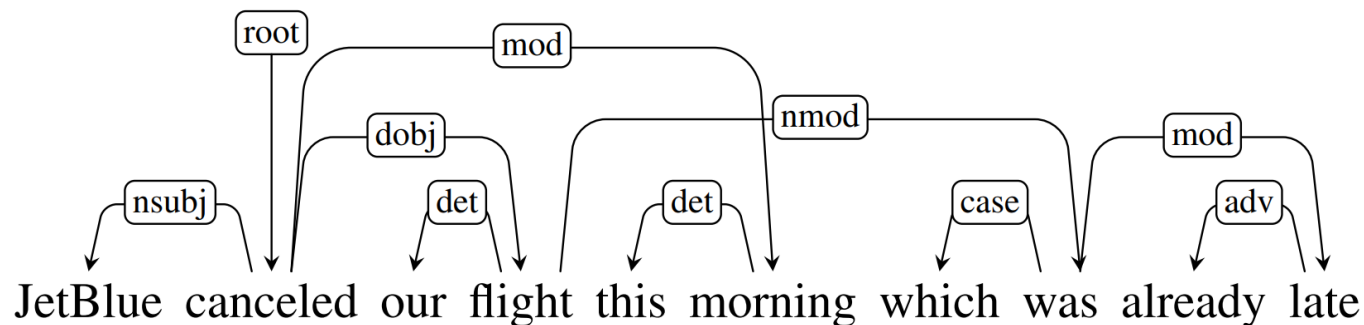
依存式语法与构成式语法的互转

例子：

依存式语法与构成式语法的互转



projective



Non-projective

- ▶ 依存式语法 (dependency grammar) 简介
 - ▶ 构成式语法的问题
 - ▶ 依存式语法基本概念
 - ▶ 依存式语法与构成式语法的互转条件
- ▶ 使用shift-reduce parser做依存式语法解析
- ▶ 评价指标
- ▶ 常用工具

使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

将句子中的单词一个个推入一个stack中,

对stack头部的两个元素进行连边, 连边后的dependent从stack删除。

最终清空stack时, 语法树就已标记完成。

使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

book	me	the	morning	flight
------	----	-----	---------	--------

依存关系输出

parser

leftArc
rightArc
shift

堆栈 (stack)

				root
--	--	--	--	------

使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

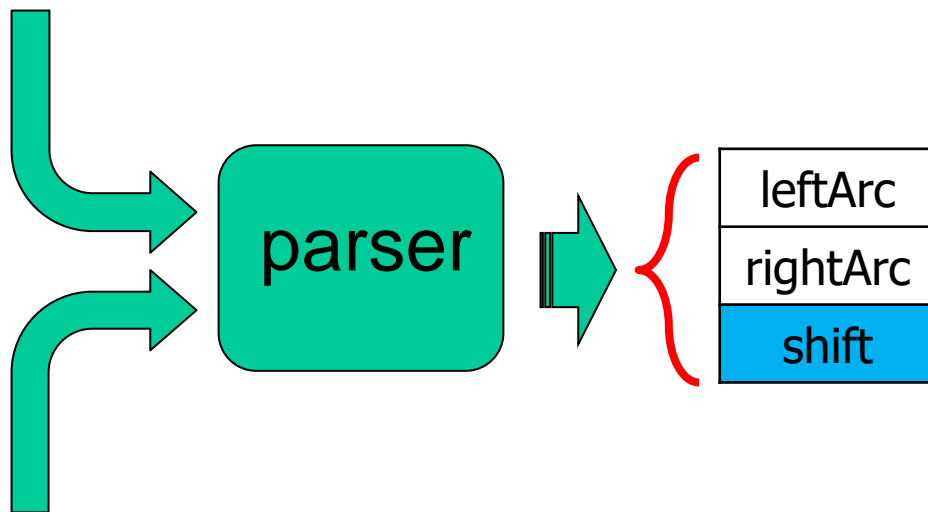
剩余词列表 (buffer)

book	me	the	morning	flight
------	----	-----	---------	--------

依存关系输出

堆栈 (stack)

				root
--	--	--	--	------



使用shift-reduce parser做依存式语法解析

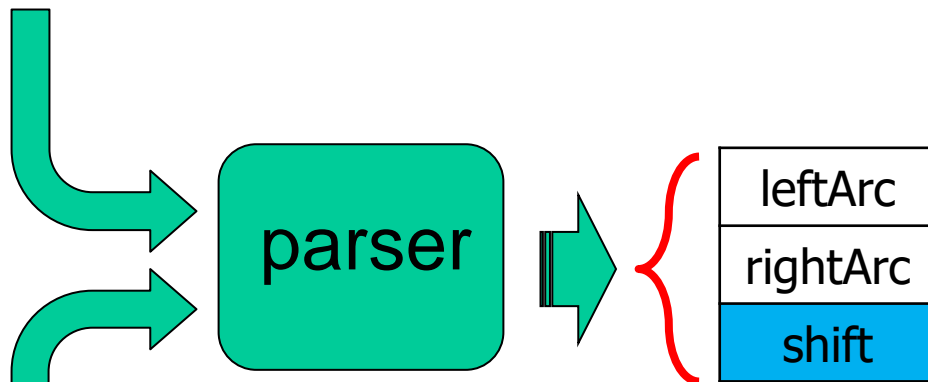
Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

me	the	morning	flight	
----	-----	---------	--------	--

依存关系输出



堆栈 (stack)

			root	book
--	--	--	------	------

使用shift-reduce parser做依存式语法解析

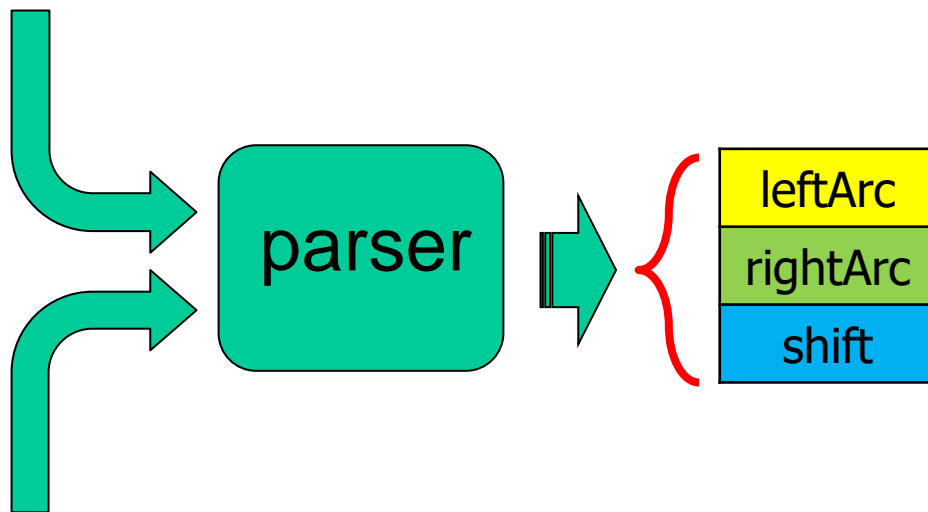
Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

me	the	morning	flight	
----	-----	---------	--------	--

依存关系输出



堆栈 (stack)

			root	book
--	--	--	------	------

使用shift-reduce parser做依存式语法解析

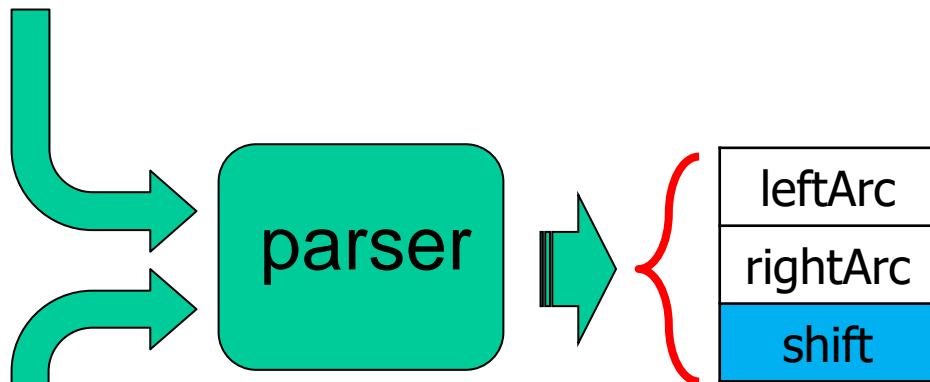
Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

me	the	morning	flight	
----	-----	---------	--------	--

依存关系输出



堆栈 (stack)

			root	book
--	--	--	------	------

使用shift-reduce parser做依存式语法解析

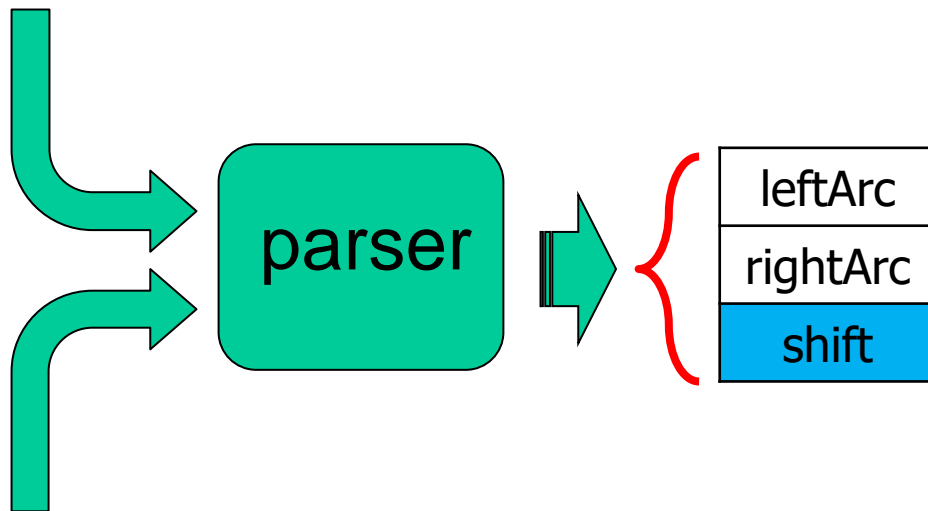
Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

the	morning	flight		
-----	---------	--------	--	--

依存关系输出



堆栈 (stack)

		root	book	me
--	--	------	------	----

使用shift-reduce parser做依存式语法解析

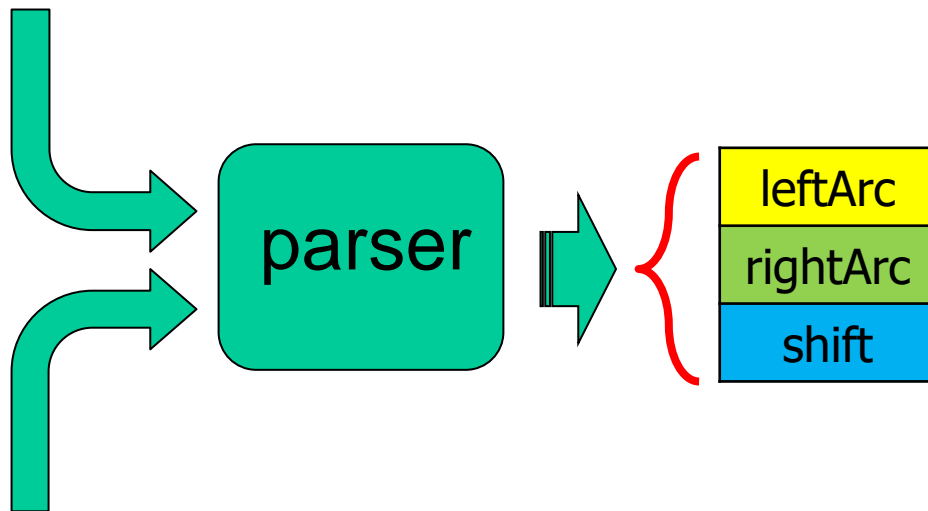
Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

the	morning	flight		
-----	---------	--------	--	--

依存关系输出



堆栈 (stack)

		root	book	me
--	--	------	------	----

使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

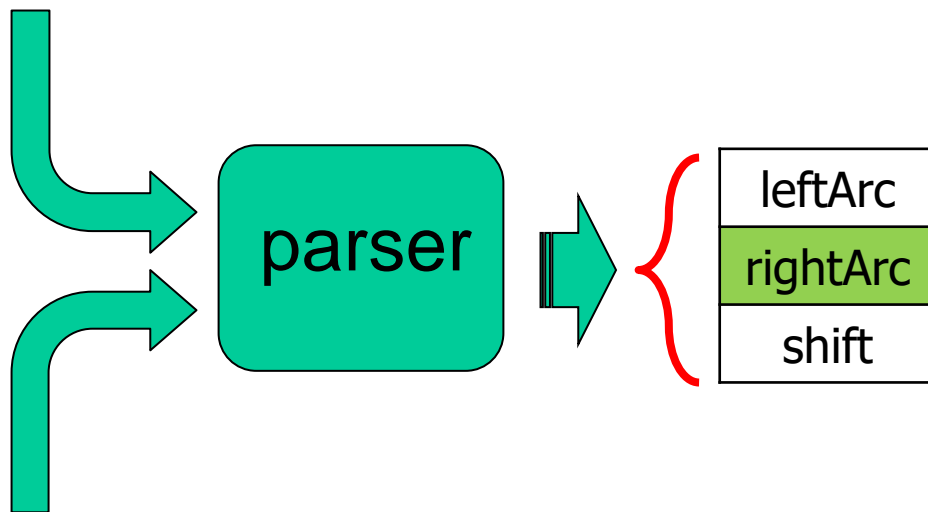
剩余词列表 (buffer)

the	morning	flight		
-----	---------	--------	--	--

依存关系输出

堆栈 (stack)

		root	book	me
--	--	------	------	----



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

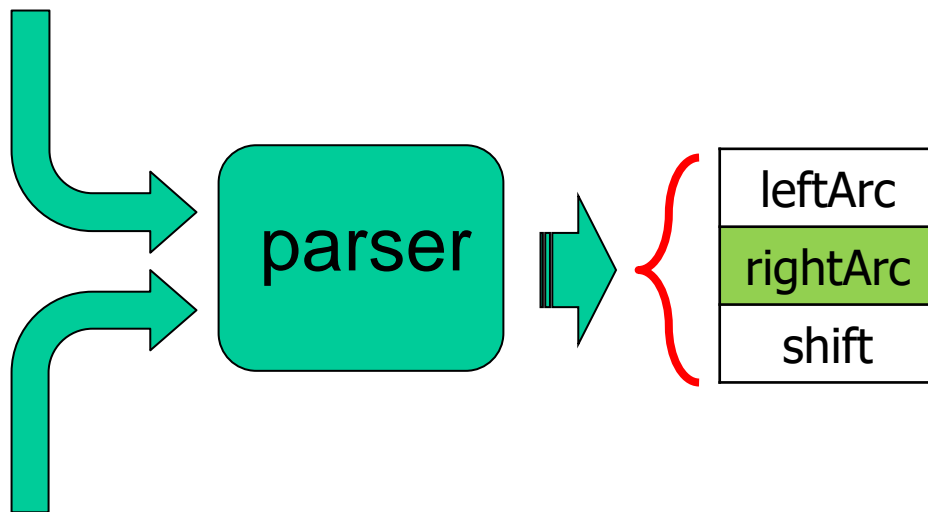
the	morning	flight		
-----	---------	--------	--	--

依存关系输出

book → **me**

堆栈 (stack)

		root	book	me
--	--	------	-------------	-----------



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

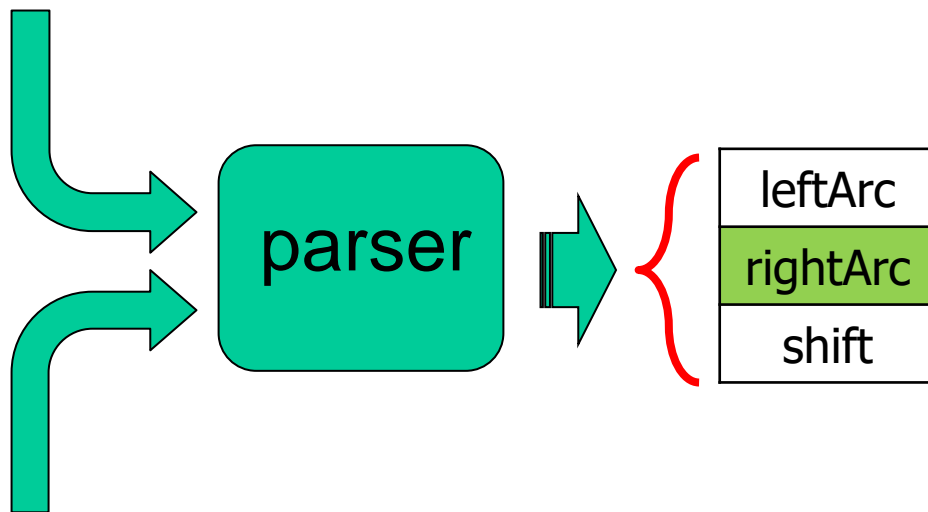
the	morning	flight		
-----	---------	--------	--	--

依存关系输出

book → **me**

堆栈 (stack)

			root	book
--	--	--	------	-------------



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

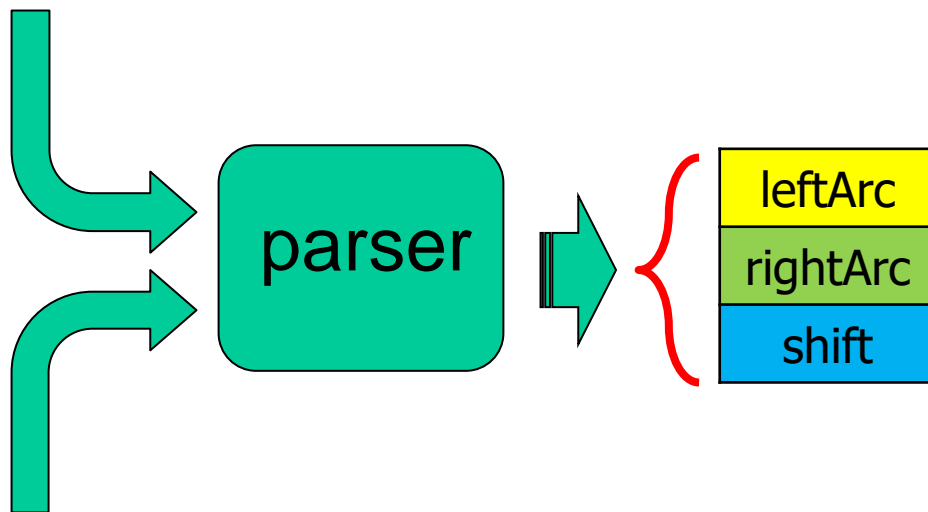
the	morning	flight		
-----	---------	--------	--	--

依存关系输出

book → **me**

堆栈 (stack)

			root	book
--	--	--	------	------



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

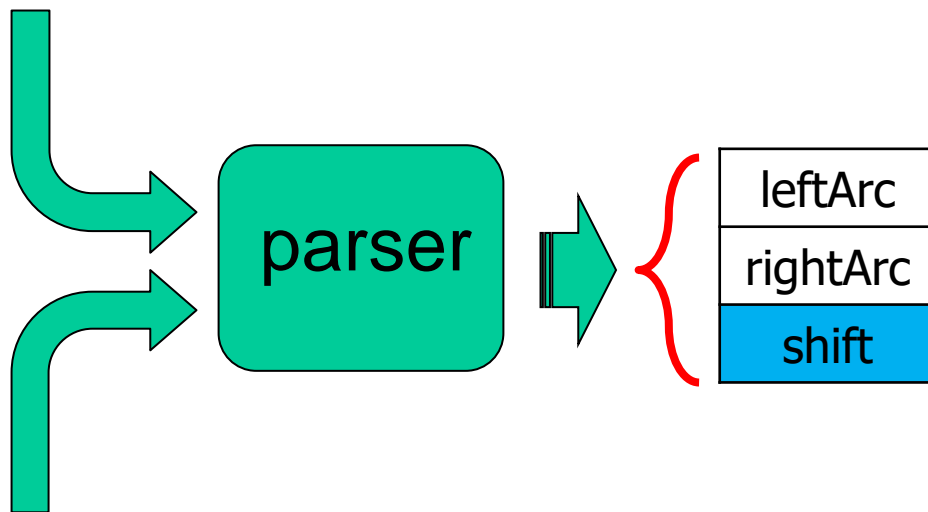
the	morning	flight		
-----	---------	--------	--	--

依存关系输出

book → **me**

堆栈 (stack)

			root	book
--	--	--	------	------



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

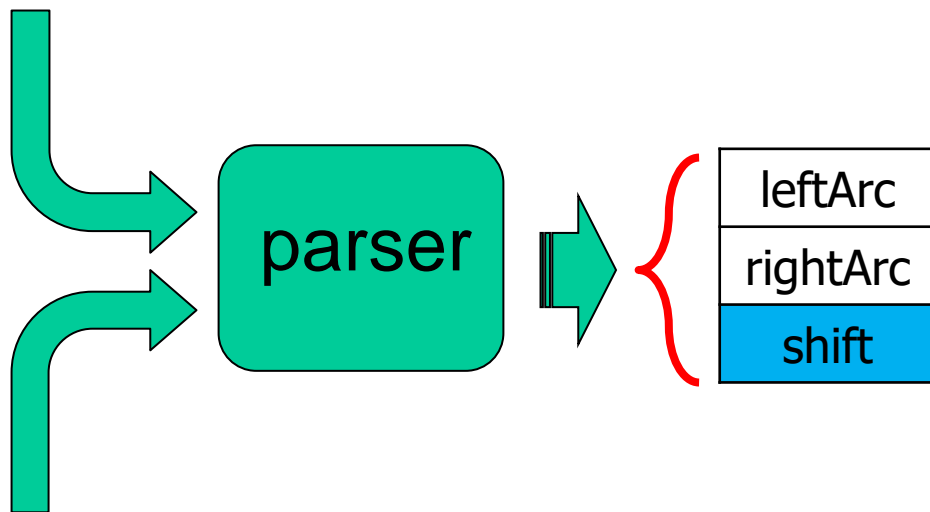
morning	flight			
---------	--------	--	--	--

依存关系输出

book → me

堆栈 (stack)

		root	book	the
--	--	------	------	-----



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

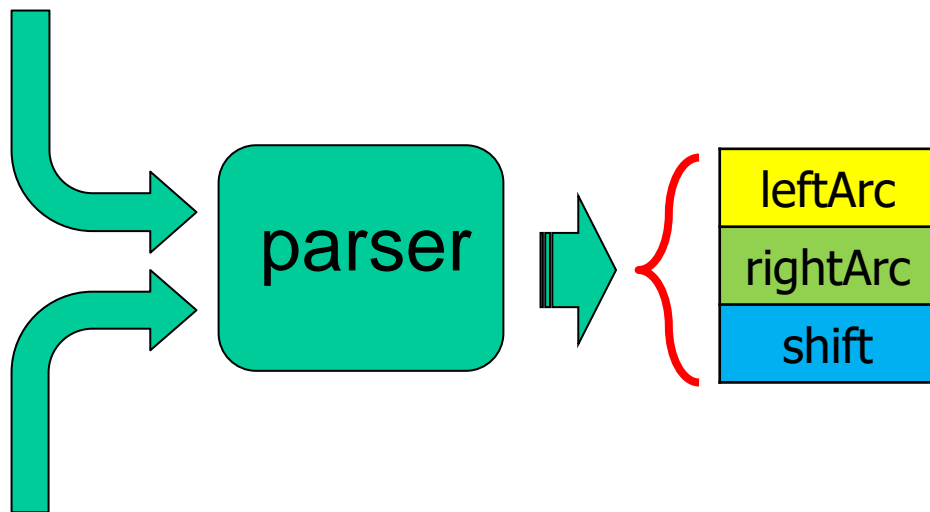
morning	flight			
---------	--------	--	--	--

依存关系输出

book → me

堆栈 (stack)

		root	book	the
--	--	------	------	-----



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

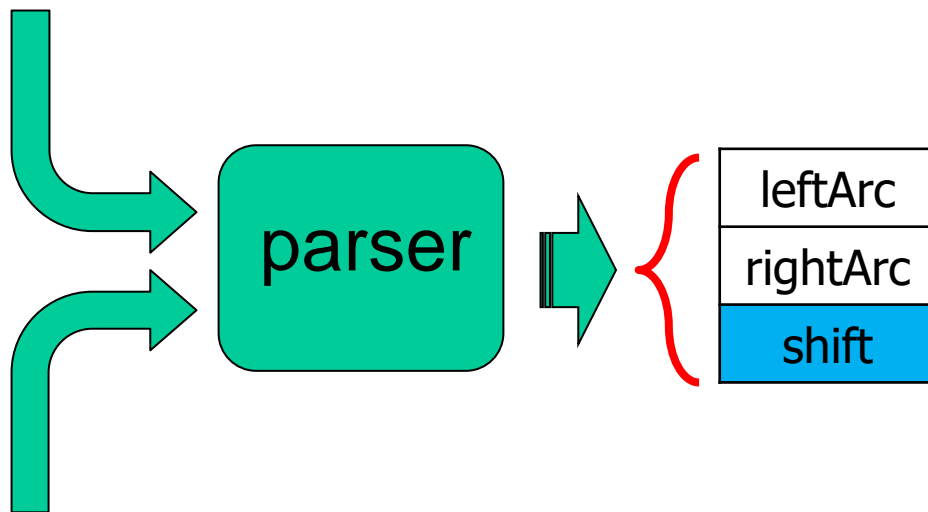
morning	flight			
---------	--------	--	--	--

依存关系输出

book → me

堆栈 (stack)

		root	book	the
--	--	------	------	-----



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

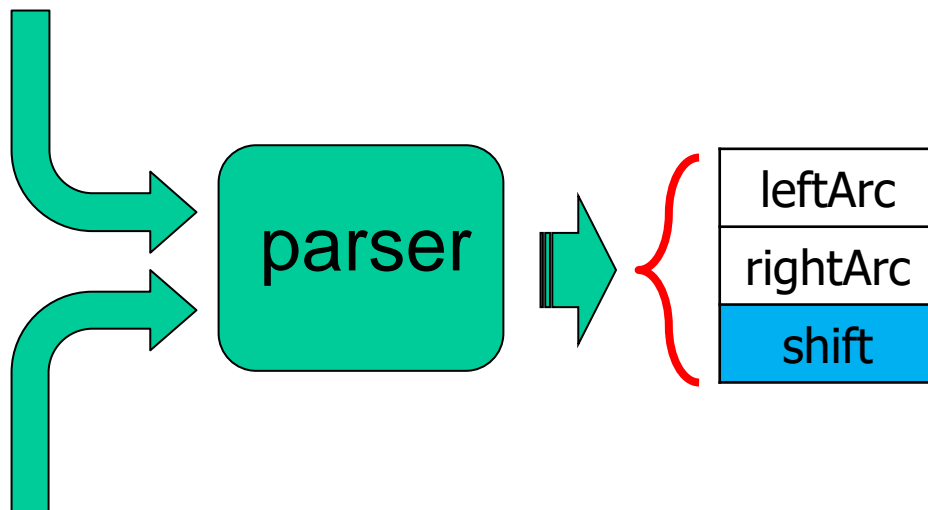
剩余词列表 (buffer)



依存关系输出

book → **me**

堆栈 (stack)



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

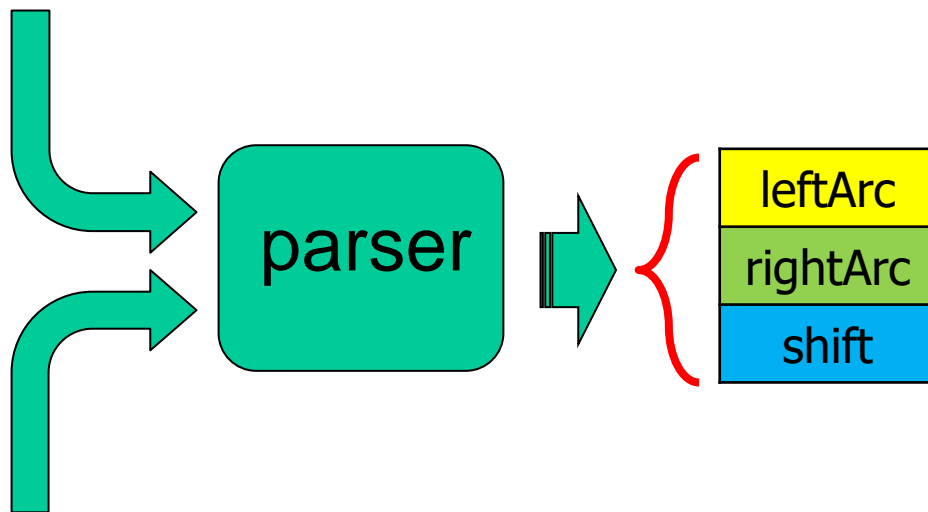
flight				
--------	--	--	--	--

依存关系输出

book → me

堆栈 (stack)

	root	book	the	morning
--	------	------	-----	---------



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

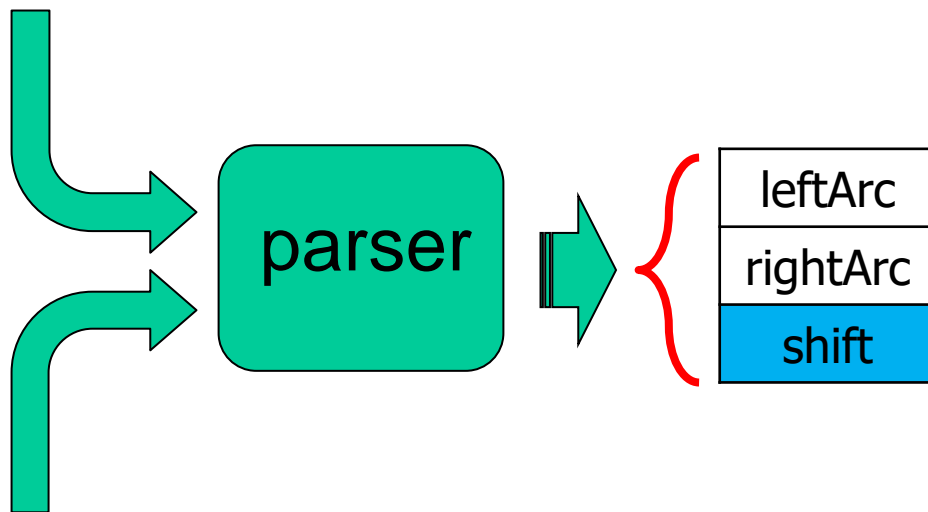
flight				
--------	--	--	--	--

依存关系输出

book → me

堆栈 (stack)

	root	book	the	morning
--	------	------	-----	---------

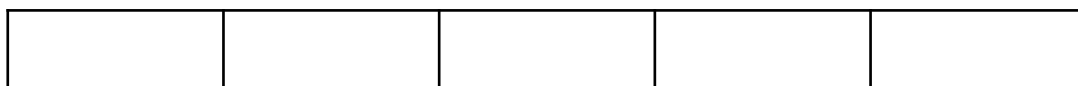


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

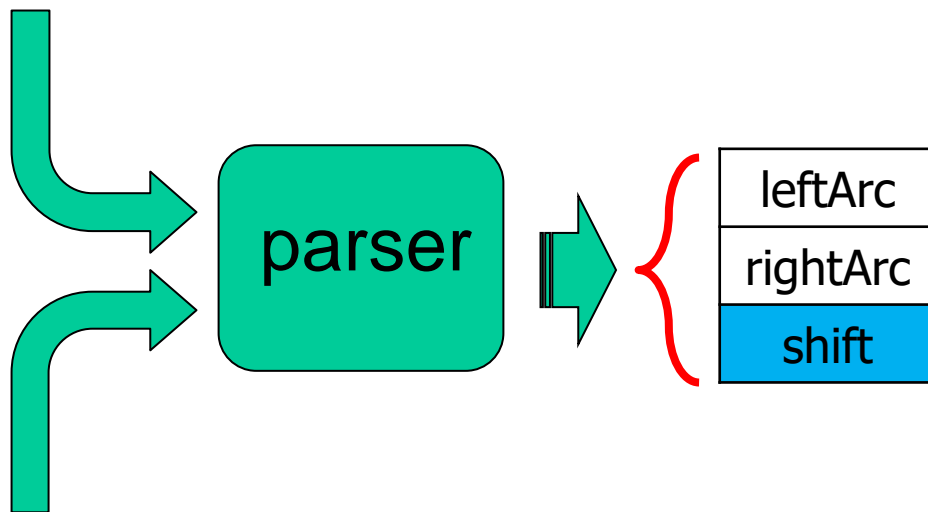
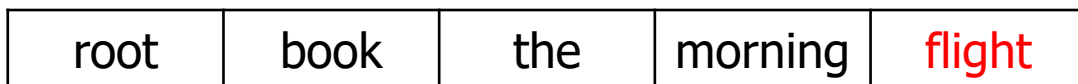
剩余词列表 (buffer)



依存关系输出

book → me

堆栈 (stack)

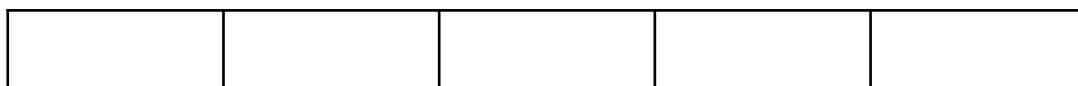


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

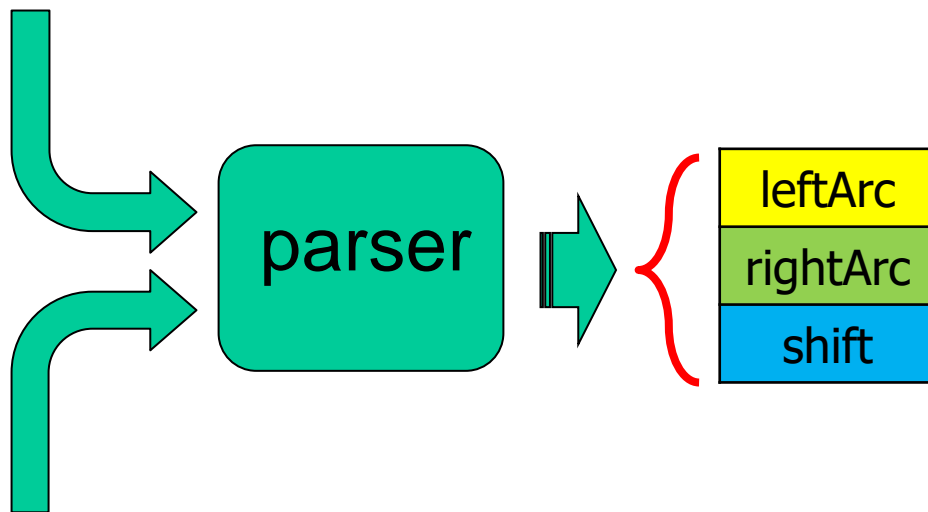
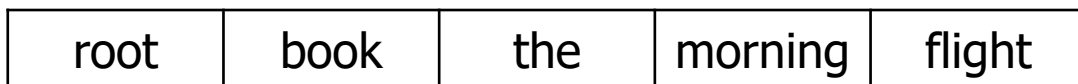
剩余词列表 (buffer)



依存关系输出

book → me

堆栈 (stack)

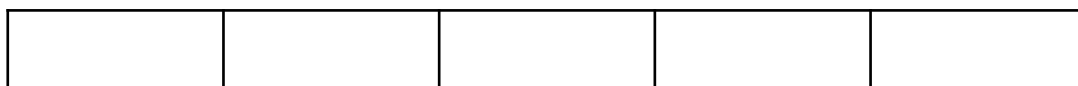


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

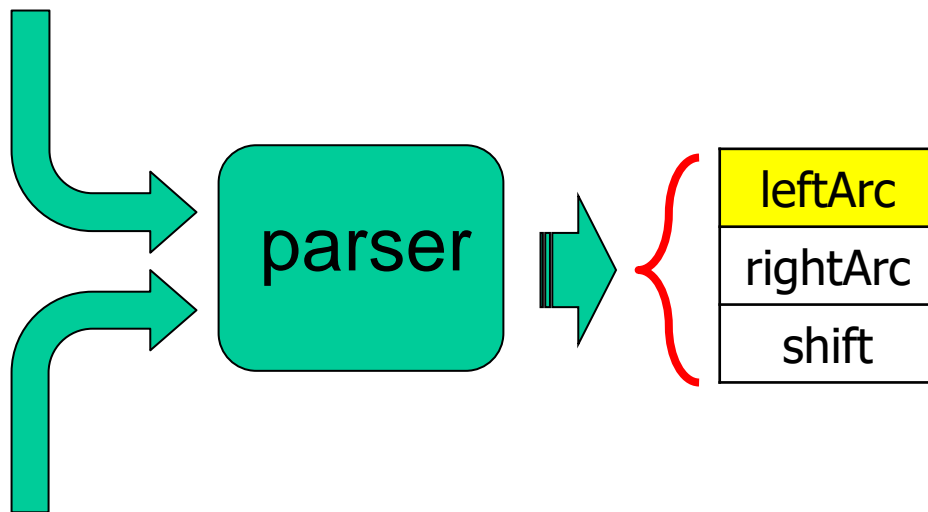
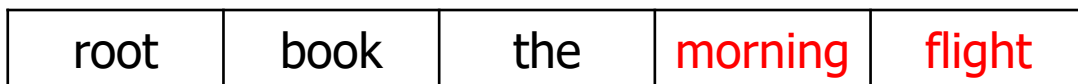


依存关系输出

book → **me**

morning ← **flight**

堆栈 (stack)

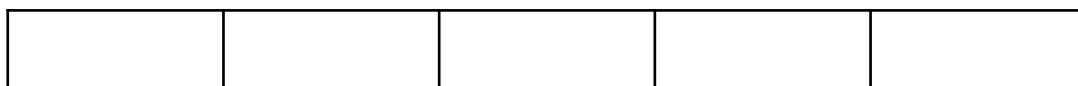


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

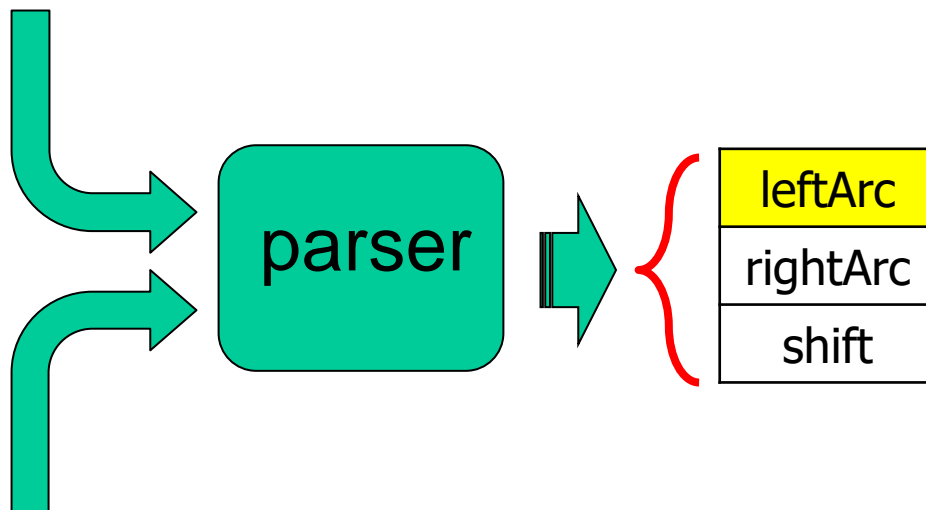


依存关系输出

book → **me**

morning ← **flight**

堆栈 (stack)

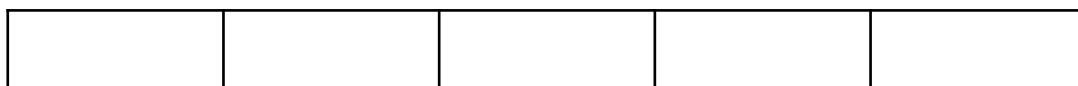


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)

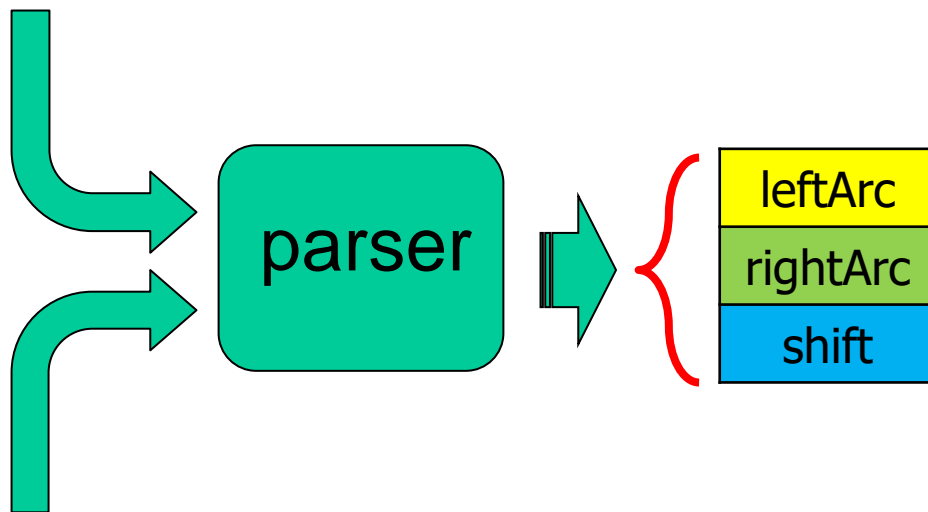


依存关系输出

book → me

morning ← flight

堆栈 (stack)

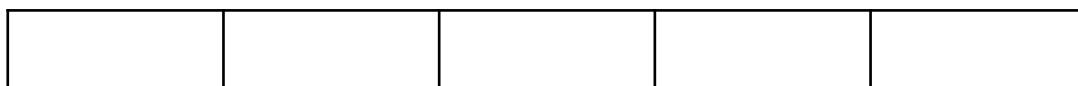


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)



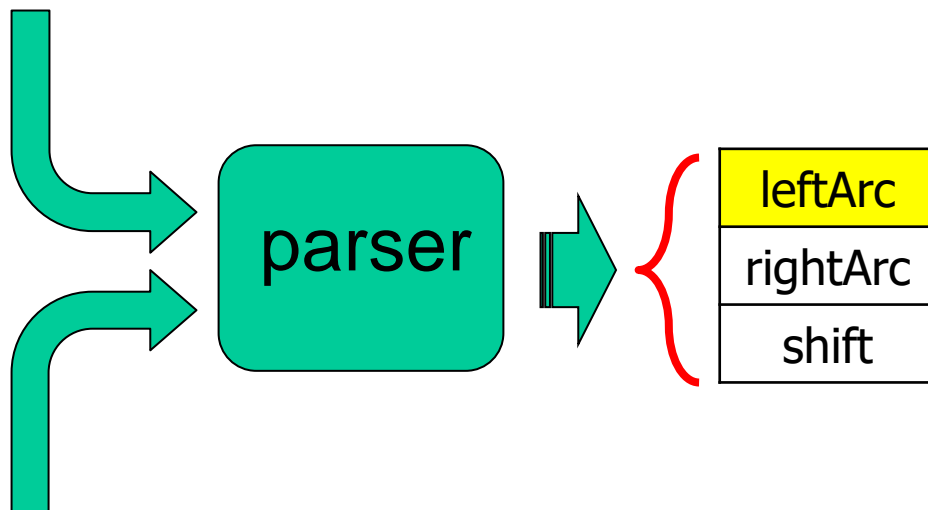
依存关系输出

book → **me**

morning ← **flight**

the ← **flight**

堆栈 (stack)

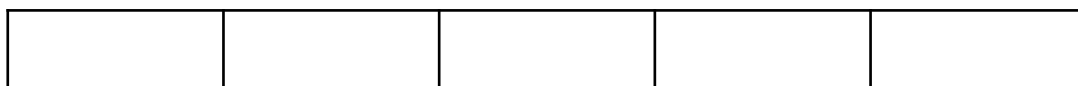


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)



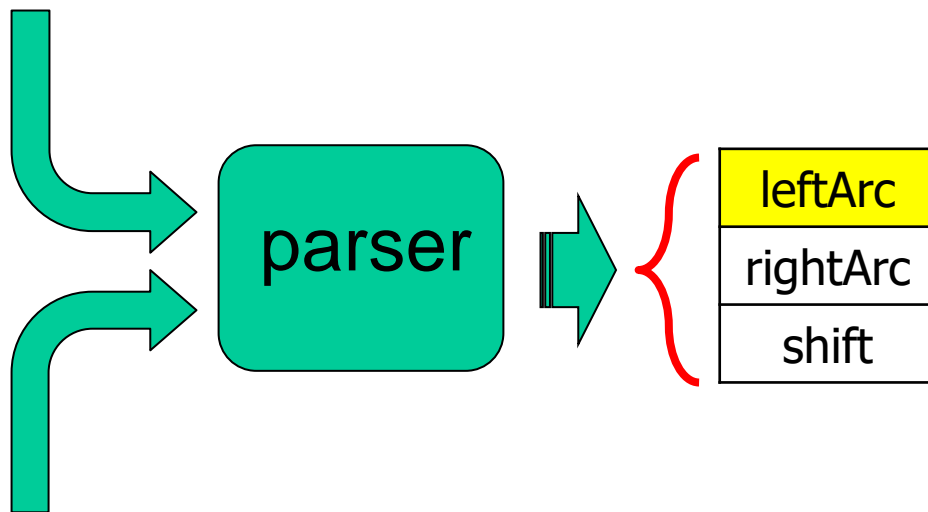
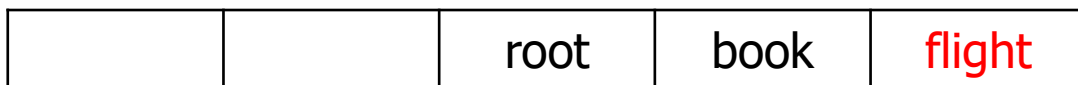
依存关系输出

book → **me**

morning ← **flight**

the ← **flight**

堆栈 (stack)

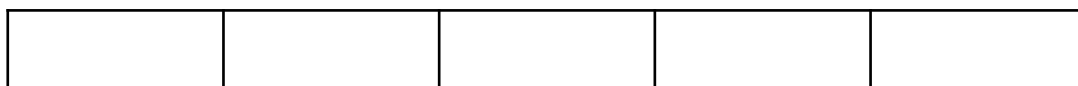


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)



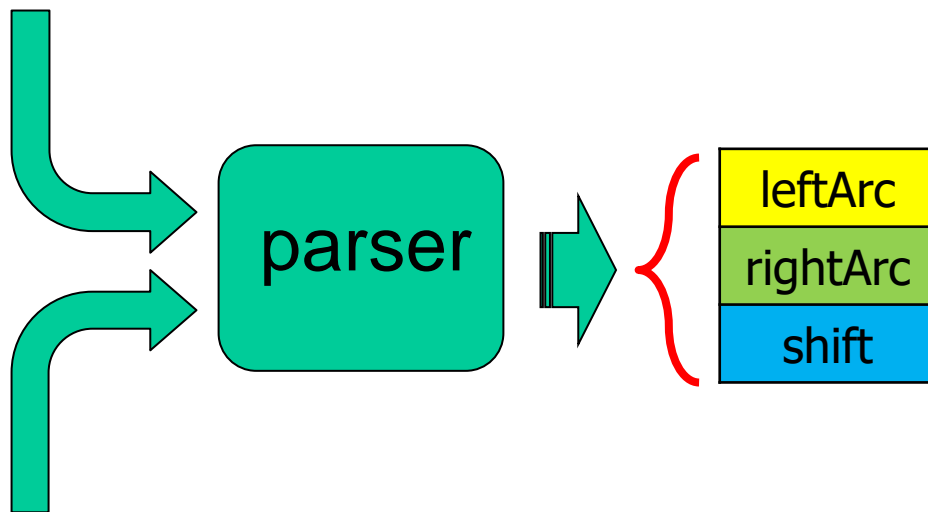
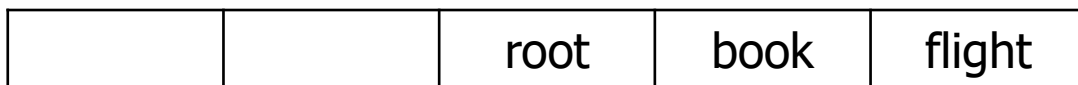
依存关系输出

book → me

morning ← flight

the ← flight

堆栈 (stack)

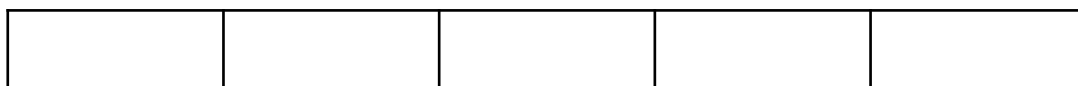


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)



依存关系输出

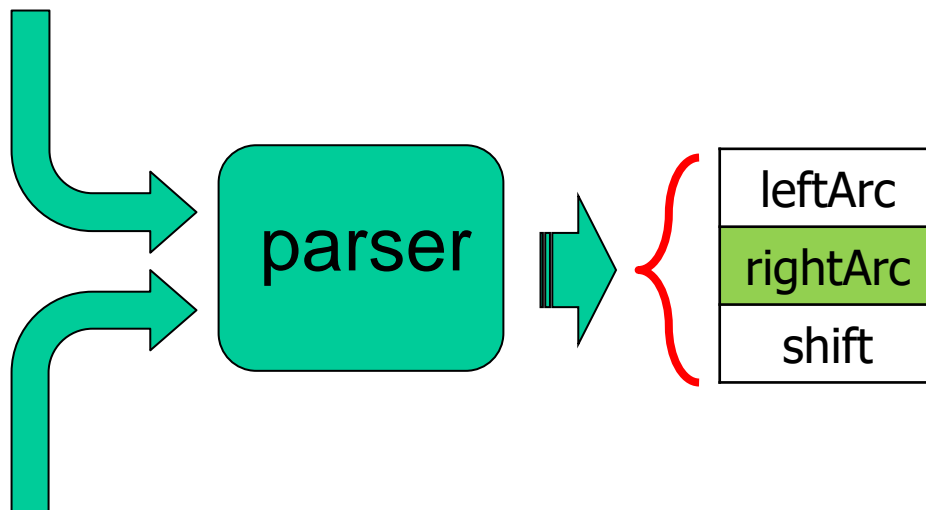
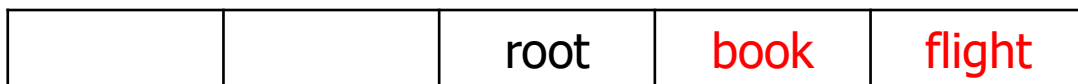
book → **me**

morning ← **flight**

the ← **flight**

book ← **flight**

堆栈 (stack)

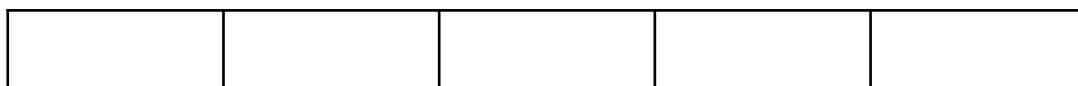


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)



依存关系输出

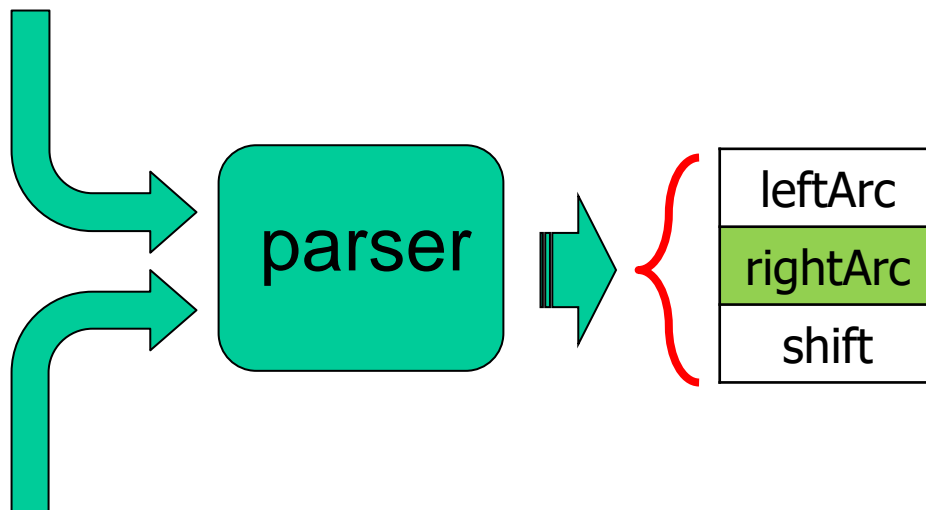
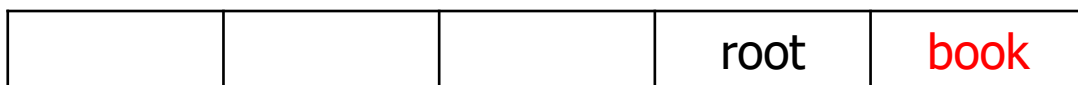
book → **me**

morning ← **flight**

the ← **flight**

book ← **flight**

堆栈 (stack)

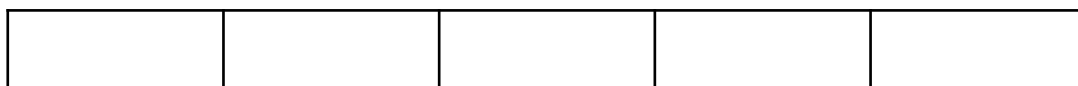


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)



依存关系输出

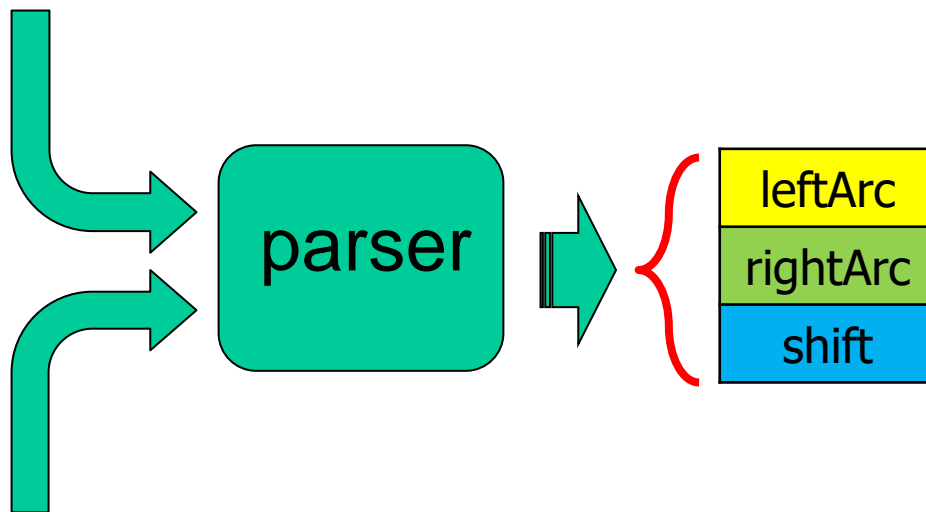
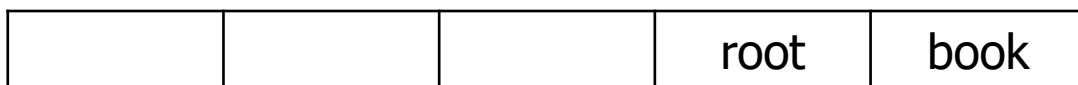
book → me

morning ← flight

the ← flight

book ← flight

堆栈 (stack)

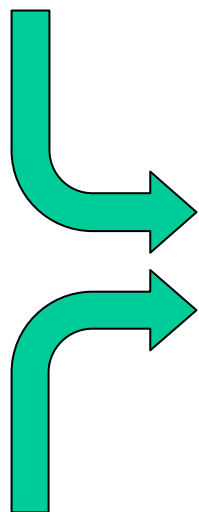
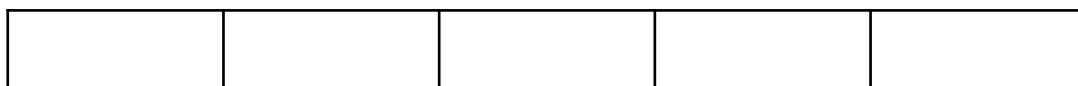


使用shift-reduce parser做依存式语法解析

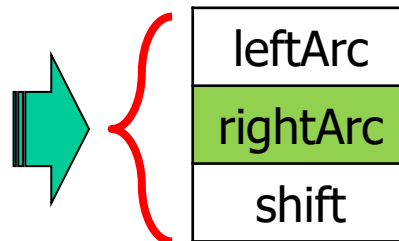
Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)



parser



堆栈 (stack)



依存关系输出

book → **me**

morning ← **flight**

the ← **flight**

book ← **flight**

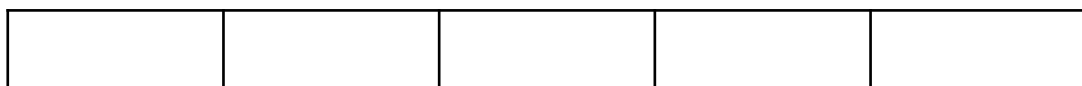
root ← **book**

使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)



依存关系输出

book → **me**

morning ← **flight**

the ← **flight**

book ← **flight**

root ← **book**

parser

leftArc

rightArc

shift

堆栈 (stack)

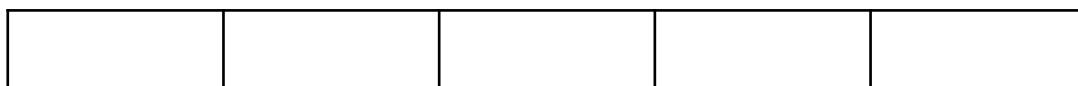


使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morning flight

剩余词列表 (buffer)



依存关系输出

book → **me**

morning ← **flight**

the ← **flight**

book ← **flight**

root ← **book**

parser

leftArc

rightArc

shift

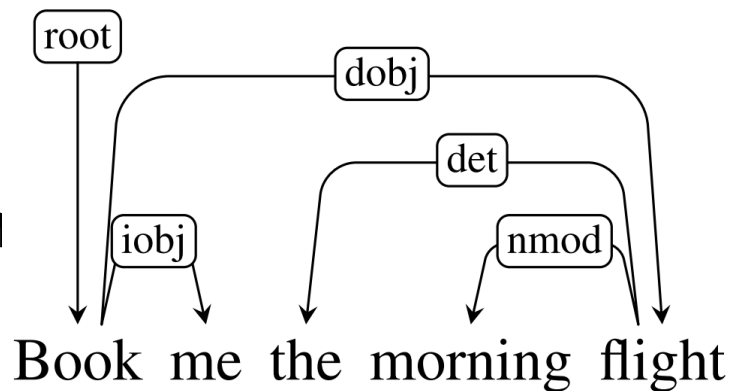
堆栈 (stack)



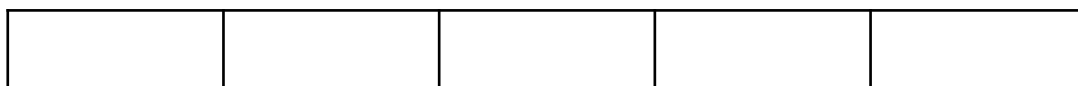
使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

Book me the morn



剩余词列表 (buffer)



依存关系输出

book → me

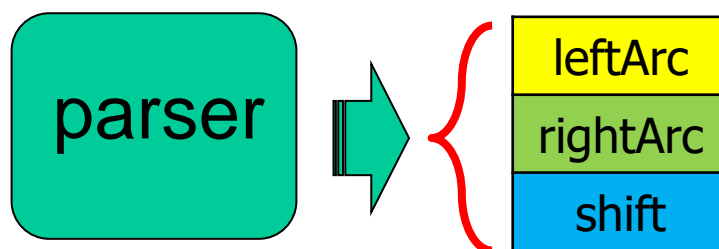
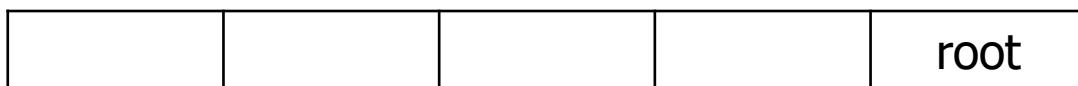
morning ← flight

the ← flight

book ← flight

root ← book

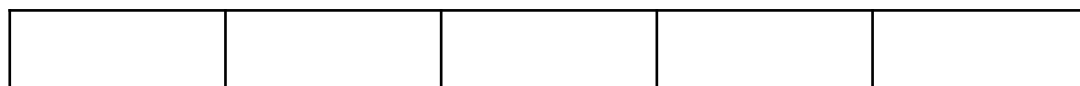
堆栈 (stack)



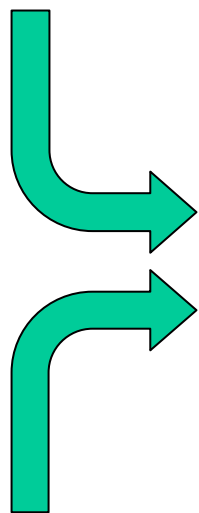
使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

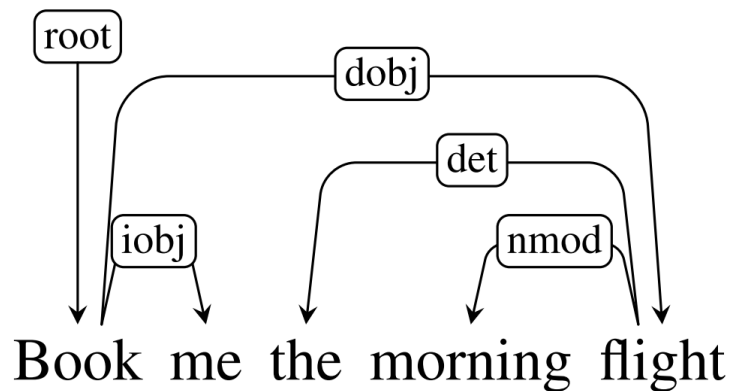
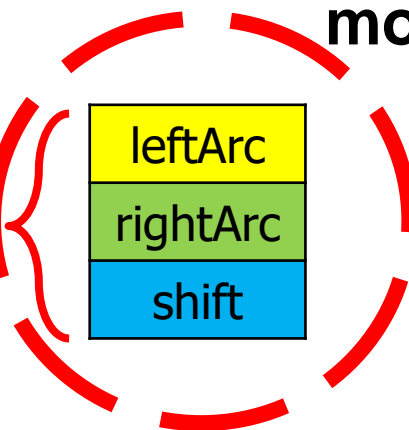
剩余词列表 (buffer)



堆栈 (stack)



parser



依存关系输出

book → **me**

morning ← **flight**

the ← **flight**

book ← **flight**

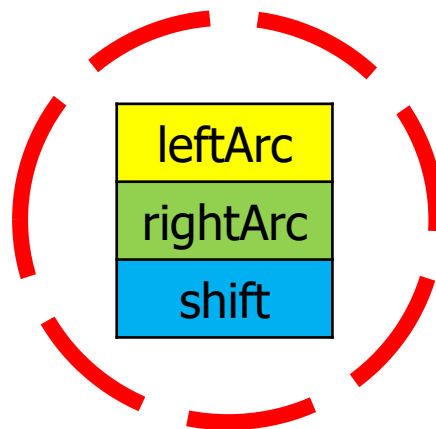
root ← **book**

使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

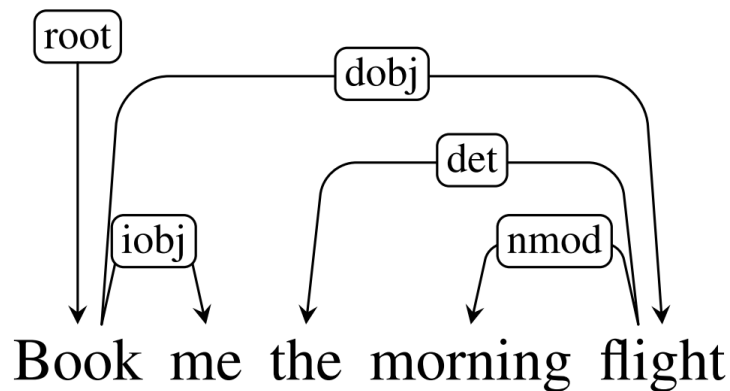
1. **leftArc** 将stack中最顶部两个词连接一条边。边的指向是从顶部第二个词指向最顶部的词。并且删除这条边的dependent。
2. **rightArc** 将stack中最顶部两个词连接一条边。边的指向是从最顶部的词指向顶部第二个词。并且删除这条边的dependent。
3. **shift** 将buffer中的下一个词推入stack中。

reduce



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式



剩余词列表 (buffer)



依存关系输出

book → **me**

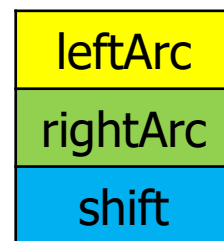
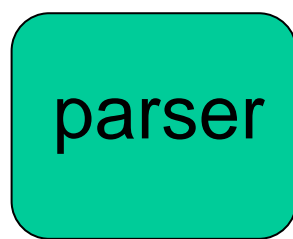
morning ← **flight**

the ← **flight**

book ← **flight**

root ← **book**

堆栈 (stack)

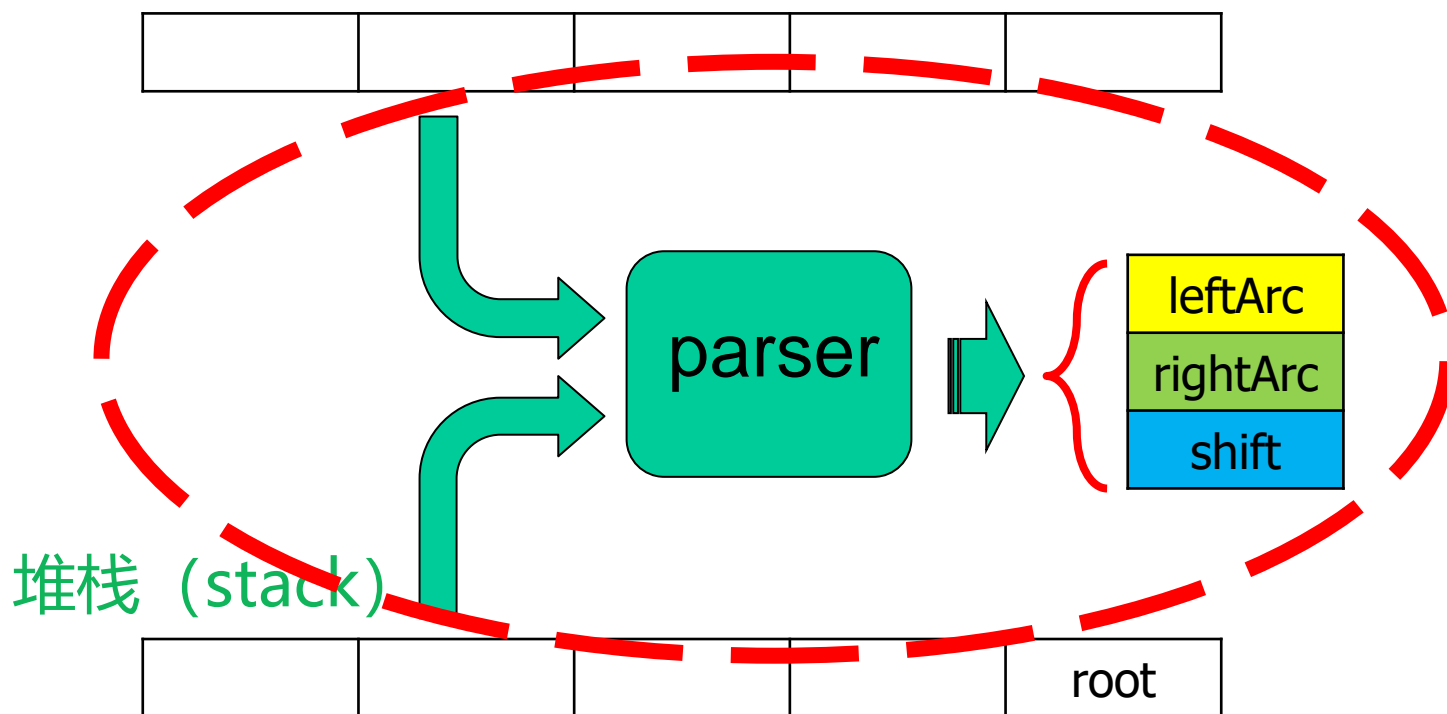


使用shift-reduce parser做依存式语法解析

Parser可看做是一个分类器

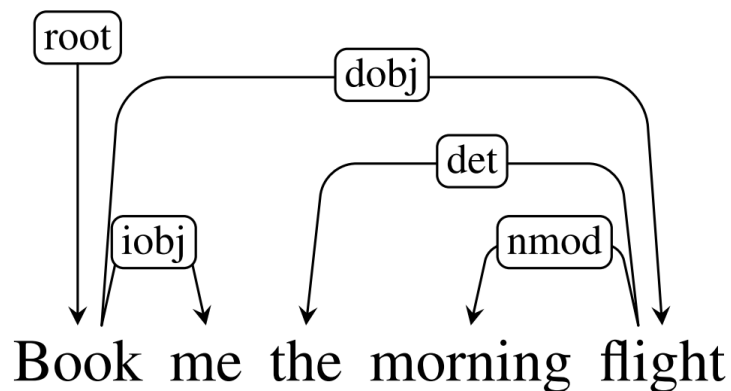
1. 读取当前时刻stack和buffer中的单词，输出三个类别的概率。
(若要预测具体arc的标签，则类别数不止三个。)
2. 分类器的feature，可按照CRF中定义feature模板的形式来定。
3. 分类器的训练，则依赖已经标记好的训练集。

剩余词列表 (buffer)



使用shift-reduce parser做依存式语法解析

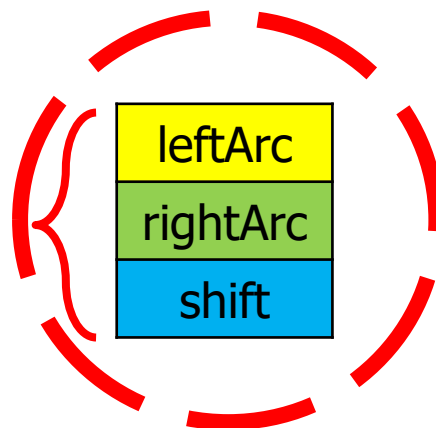
Step	Stack	Word List	Action	Relation Added
0	[root]	[book, me, the, morning, flight]	SHIFT	
1	[root, book]	[me, the, morning, flight]	SHIFT	
2	[root, book, me]	[the, morning, flight]	RIGHTARC	(book → me)
3	[root, book]	[the, morning, flight]	SHIFT	
4	[root, book, the]	[morning, flight]	SHIFT	
5	[root, book, the, morning]	[flight]	SHIFT	
6	[root, book, the, morning, flight]	[]	LEFTARC	(morning ← flight)
7	[root, book, the, flight]	[]	LEFTARC	(the ← flight)
8	[root, book, flight]	[]	RIGHTARC	(book → flight)
9	[root, book]	[]	RIGHTARC	(root → book)
10	[root]	[]	Done	



使用shift-reduce parser做依存式语法解析

Shift-reduce parser的解析形式

1. **leftArc** 只要分类器愿意, 即可选择该操作。
2. **rightArc** 存在限制条件: 最顶部的词的所有dependents都已经被确定。
3. **shift** 只要分类器愿意, 即可选择该操作。



使用shift-reduce parser做依存式语法解析

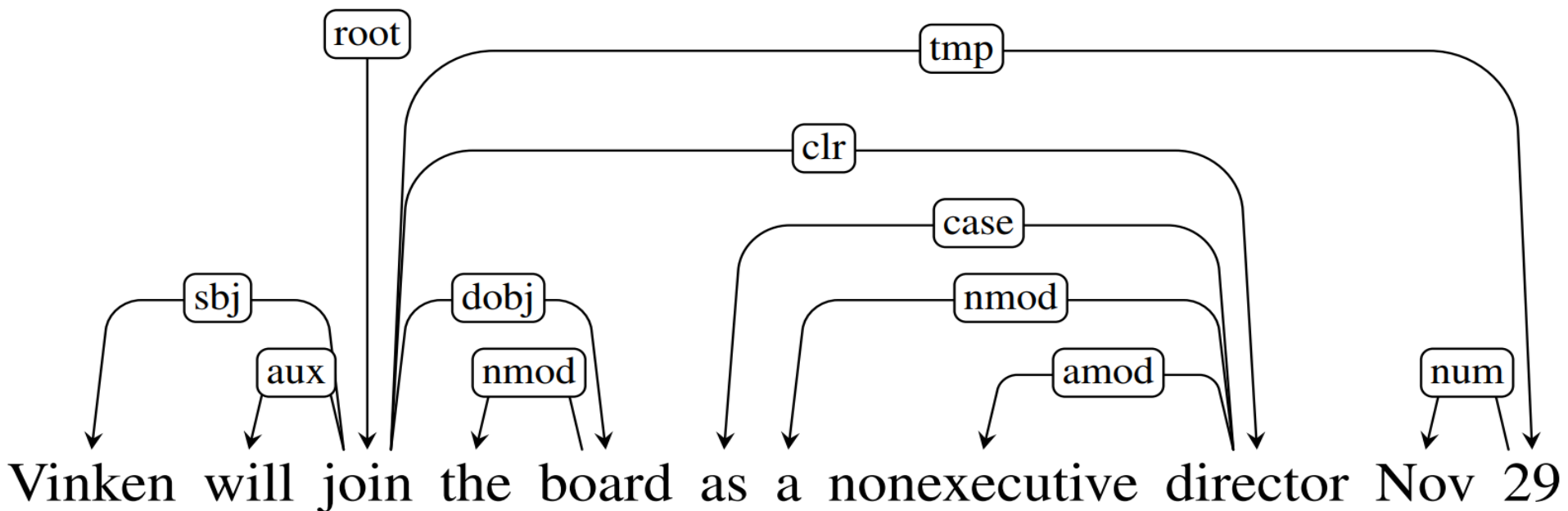
Step	Stack	Word List	Action	Relation Added
0	[root]	[book, me, the, morning, flight]	SHIFT	(book → me)
1	[root, book]	[me, the, morning, flight]	SHIFT	
2	[root, book, me]	[the, morning, flight]	RIGHTARC	
3	[root, book]	[the, morning, flight]	SHIFT	
4	[root, book, the]	[morning, flight]	SHIFT	
5	[root, book, the, morning]	[flight]	SHIFT	(morning ← flight)
6	[root, book, the, morning, flight]	[]	LEFTARC	
7	[root, book, the, flight]	[]	LEFTARC	
8	[root, book, flight]	[]	RIGHTARC	
9	[root, book]	[]	RIGHTARC	
10	[root]	[]	Done	(root → book)

思考题:

1. 与之前介绍的CKY parsing相比, CKY通过动态规划保证找到的parsing中一定包含全局最优解。
该算法是否能保证找到全局最优解?
2. 如何提升算法的搜索能力?

- ▶ 依存式语法 (dependency grammar) 简介
 - ▶ 构成式语法的问题
 - ▶ 依存式语法基本概念
 - ▶ 依存式语法与构成式语法的互转条件
- ▶ 使用shift-reduce parser做依存式语法解析
- ▶ 评价指标
- ▶ 常用工具

评价指标



每个单词有且仅有一次被作为dependent!

只要去数每个单词作为dependent时，他的head有没找对即可。

UAS (unlabeled attachment score):
每个单词的head被找对的比例。

LAS (labeled attachment score)

每个单词的head被找对，并且边的标签也对的比例。

- ▶ 依存式语法 (dependency grammar) 简介
 - ▶ 构成式语法的问题
 - ▶ 依存式语法基本概念
 - ▶ 依存式语法与构成式语法的互转条件
- ▶ 使用shift-reduce parser做依存式语法解析
- ▶ 评价指标
- ▶ 常用工具

在线语法解析demo:

<https://explosion.ai/demos/displacy> (Berkeley Neural Parser)

<http://stanza.run/> , <https://corenlp.run/> (Stanford coreNLP demo)

http://bionlp-www.utu.fi/parser_demo/ (英语和芬兰语的parser)

<https://hanlp.hankcs.com/> (面向中文的Parser demo)

repo和模型:

Model	UAS	LAS	Paper / Source
Label Attention Layer + HPSG + XLNet (Mrini et al., 2019)	97.42	96.26	Rethinking Self-Attention: Towards Interpretability for Neural Parsing
ACE + fine-tune (Wang et al., 2020)	97.20	95.80	Automated Concatenation of Embeddings for Structured Prediction
HPSG Parser (Joint) + XLNet (Zhou et al, 2020)	97.20	95.72	Head-Driven Phrase Structure Grammar Parsing on Penn Treebank
Second-Order MFVI + BERT (Wang et al., 2020)	96.91	95.34	Second-Order Neural Dependency Parsing with Message Passing and End-to-End Training
CVT + Multi-Task (Clark et al., 2018)	96.61	95.02	Semi-Supervised Sequence Modeling with Cross-View Training
CRF Parser (Zhang et al., 2020)	96.14	94.49	Efficient Second-Order TreeCRF for Neural Dependency Parsing