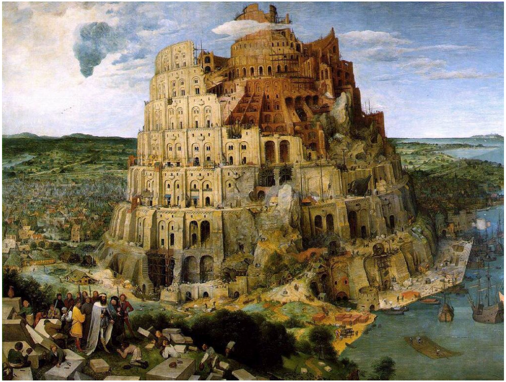
An Incremental Syntactic Language Model for Statistical Phrase-based Machine Translation

Lane Schwartz

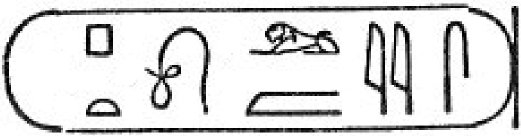
Air Force Research Lab
lane.schwartz@wpafb.af.mil

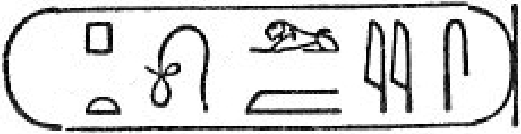
Akeqiinga malighqutaqnalunga tuqlughaasiiniun America-m ama nunganun nekevghaviganun ataasiq nunaghllak asingani Kiyaghneghem, ilemngalghii ilakutelleq ama ataasiighngalghhi tamaghhaanun.

7,000,000,000 people



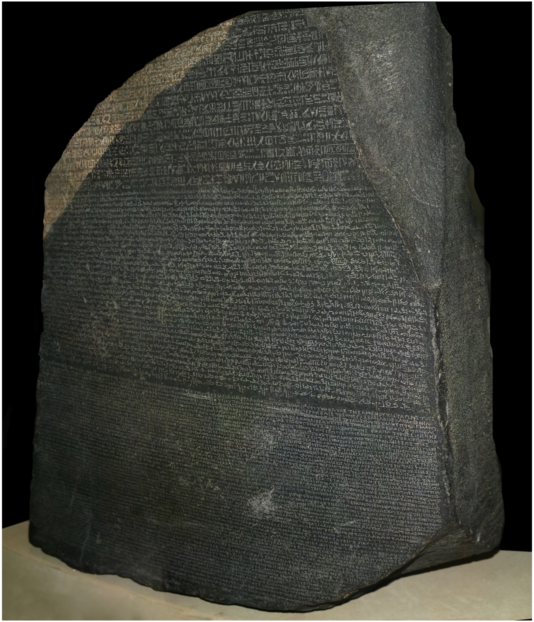
8 languages with at least 100 million speakers
85 languages with at least 10 million speakers
389 languages with at least 1 million speakers
6632 languages with at least 1 speaker





Πτολεμαις

Κλεοπατρα



頭抬	APPETIZERS	Lunch	Dinner
上海春卷	Crispy Spring Roll (2)	3	3
素菜春卷	Vegetable Spring Roll (2)	3	3
蒸餃	Steamed Meat Dumplings (6)	5	6
鍋貼	Pan Fried Meat Dumplings (6)	5	6
日式雞串	Teriyaki Chicken Sticks (4)	6	7
蟹肉雲吞	Crab Rangoon (6)	5	6
燒排骨	Bar-B-Q Spareribs (4)		7
寶寶盤	Pu Pu Tray (For 2)		13
涼拌海蜇	Cold Jelly Fish		7
佛山燻蹄	Slicked Boneless Pig's Knuckle (Cold)		8
海蜇拌火鴨絲	Jelly Fish w. Shredded Roast Duck		12
五香牛展	Marinated Sliced Beef (Cold)		8
酥炸大腸	Crispy Pig's Intestine		8
椒鹽白飯魚	Crispy Silver Fish w. Spicy Salt & Pepper		10

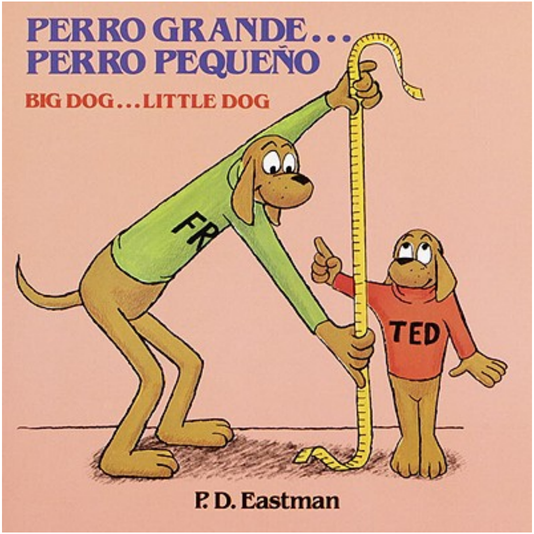
湯	SOUP	Lunch	Dinner
酸辣湯	Hot & Sour Soup	2	2
雲吞湯	Wonton Soup	2	2
蛋花湯	Egg Drop Soup	2	2
素菜湯	Vegetable Soup (For 2)	2	6
雞茸粟米羹	Chicken & Corn Chowder (For 2)		6
海鮮酸辣湯	Seafood H & S Chowder (For 2)		9
海皇豆腐羹	Seafood Tofu Chowder (For 2)		9
韭黃瑤柱羹	Dried Scallop w. Yellow Chives Soup (For 2)		13
蟹肉魚肚羹	Fish Maw w. Crab Meat Soup (For 2)		12
香茜皮蛋魚片湯	Fish Filet & Thousand Egg w. Cilantro Soup (For 2)		10
勝瓜花蜆湯	Baby Clam w. Young Squash Soup (For 2)		10
[illegible]	Pork & Tofu w. Watercress Soup (For 2)		8
酸菜肚片湯	Sour Cabbage w. Pork Tripe Soup (For 2)		8

煲	CASSEROLE	
八珍豆腐煲	Assorted Meat & Seafood w. Tofu	15
[illegible]	Tofu & Chicken w. Salted Fish	13
蘿蔔牛腩煲	Beef Brisket w. Turnips	13
鱈魚豆腐煲	Canada Cod Fish w. Tofu	18
辣醬肉碎茄煲	Eggplant w. Spicy Ground Pork	13
沙爹粉絲牛仔骨煲	Short Rib & Vermicelli in Sate Sauce	16
勝瓜沙蜆煲	Baby Clams w. Young Squash	16
姜葱生蠔煲	Oyster w. Ginger & Scallion	15
雙冬田雞煲	Frog w. Mushroom & Bamboo Shoot	16
啫啫田雞煲	Frog w. Ginger & Scallion	16
枝竹羊腩煲	Lamb w. Dried Bean Curd	16
節瓜粉絲蝦米煲	Dried Shrimp & Squash w. Vermicelli	12
姜葱雞片煲	Sliced Chicken w. Ginger & Scallion	13

雞	CHICKEN	Lunch	Dinner
左宗雞	General Tso's Chicken		13
芝麻雞	Sesame Chicken	9	13
西檸雞	Chicken w. Lemon Sauce	9	13
陳皮雞	Orange Chicken	9	13
甜酸雞	Sweet & Sour Chicken		12
蘑菇雞	Moo Goo Gai Pan		12
湖南雞	Chicken, Hunan Style		12
腰果雞	Chicken w. Cashew Nuts		12
宮保雞	Kung Pao Chicken		12
豉汁雞	Chicken in Black Bean Sauce		12
咖喱雞	Chicken in Curry Sauce		12
一品蒜香雞	Crispy Fried Chicken w. Garlic	(Half) 13	(Whole) 24
明爐燒鴨	Crispy Roast Duck	(Half) 14	(Whole) 27
北京片皮鴨	Peking Duck Served w. 4 or 8 pancakes, onion & plum sauce.	(Half) 17	(Whole) 32
芥蘭雞	Chicken w. Broccoli	8	
四川雞	Szechuan Chicken	8	
魚香雞	Chicken in Garlic Sauce	8	

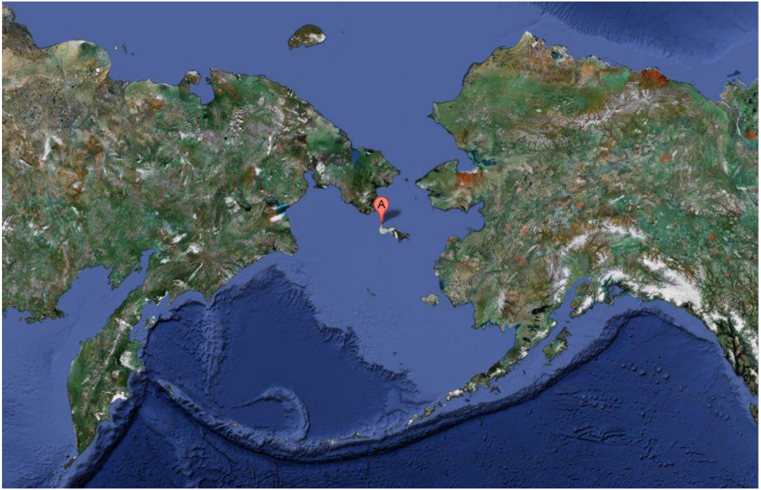
海鮮	SEAFOOD	Lunch	Dinner
西檸汁魚柳	Fish Filet w. Lemon Sauce	12	14
左宗蝦	General Tso's Shrimp		15
甜酸蝦	Sweet & Sour Shrimp		14
蝦龍糊	Shrimp w. Lobster Sauce		15
湖南蝦	Shrimp, Hunan Style		15
四川蝦	Shrimp, Szechuan Style		15
素菜帶子	Scallop w. Mixed Vegetable	13	15
腰果蝦仁	Shrimp w. Cashew Nut		13
宮保雙丁	Kung Pao Shrimp & Chicken	12	13
宮保雙鮮	Kung Pao Shrimp & Scallop		17
鮮魷酸菜	Sour Cabbage w. Squid		13
玉蘭干貝	Scallop Stir-Fried w. Chinese Broccoli		15
時菜斑球	Fish Filet Stir-Fried w. Chinese Vegetable	12	15
水煮魚	Fish Filet in Super Spicy Sauce, Szechuan Style		15
五更海鮮	Seafood Combo in Chef's Spicy Sauce		17
三杯小卷	Three Cup Sauce Squid		15
炒三鮮	Triple Delight	14	

肉	PORK	Lunch	Dinner
甜酸肉	Sweet & Sour Pork	8	12
回鍋肉	Double Cooked Pork		12
魚香肉絲	Pork in Hot Garlic Sauce		12
葱爆肉	Mongolian Pork	8	12
叉燒雪豆	Roast Pork w. Snow Peas		12
尖椒炒肚絲	Pork Intestine w. Hot Pepper		13
香幹肉絲	Shredded Pork w. Shredded Dried Tofu		12
木須肉	Moo Shi Pork (4 Pancakes)	9	13
叉燒素菜	Roast Pork w. Mixed Vegetable	8	
辣椒肉絲	Shredded Pork w. Hot Pepper	8	

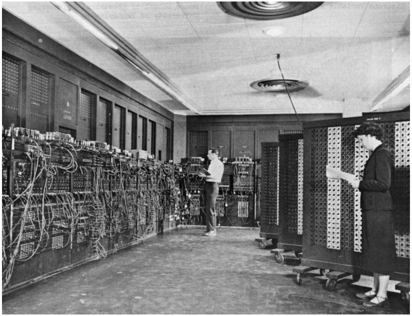

PERRO GRANDE...
PERRO PEQUEÑO
BIG DOG...LITTLE DOG
FR
TED
P. D. Eastman

O Canada!
Our home and native land!
True patriot love
    in all thy sons command.
With glowing hearts we see thee rise,
The True North strong and free!

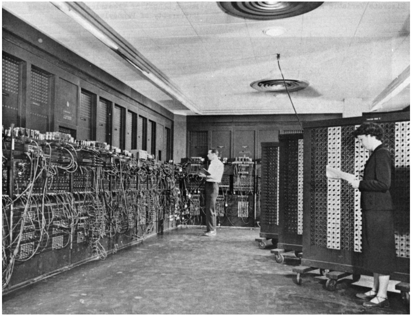
Ô Canada!
Terre de nos aïeux,
Ton front est ceint
    de fleurons glorieux!
Car ton bras sait porter l'épée,
Il sait porter la croix!

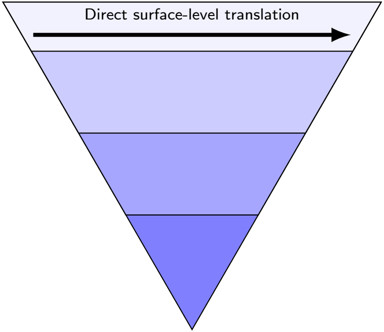

A

Rule-Based Machine Translation

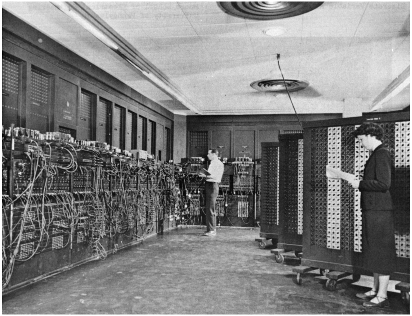


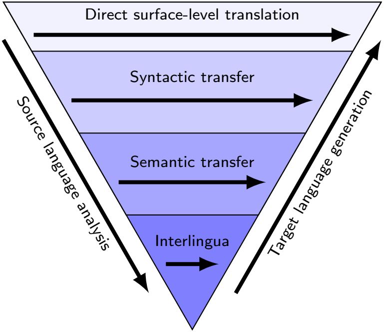
Rule-Based Machine Translation



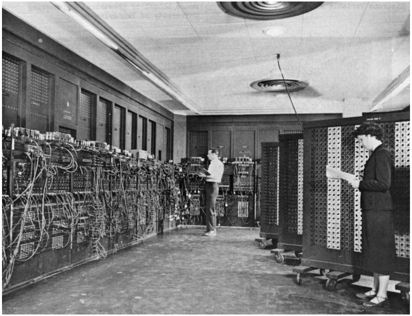


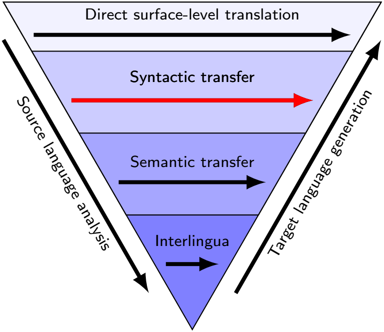
Rule-Based Machine Translation





Rule-Based Machine Translation





Machine Translation Insights — Warren Weaver

"One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography."

Machine Translation Insights — Warren Weaver

"When I look at an article in Russian, I say 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

Statistical Machine Translation

- Noisy Channel Model
  $$\hat{\boldsymbol{e}} = \underset{\boldsymbol{e}}{\operatorname{argmax}} \, \mathsf{P}(\boldsymbol{f} \mid \boldsymbol{e}) \mathsf{P}(\boldsymbol{e})$$

Statistical Machine Translation

- Noisy Channel Model
  $$\hat{\boldsymbol{e}} = \underset{\boldsymbol{e}}{\operatorname{argmax}} \, \mathsf{P}(\boldsymbol{f} \mid \boldsymbol{e})\mathsf{P}(\boldsymbol{e})$$

  Translation Model

- Word-Based Translation
  Brown *et al*. (1988,1993)

Statistical Machine Translation

- Noisy Channel Model
  $$\hat{\boldsymbol{e}} = \underset{\boldsymbol{e}}{\operatorname{argmax}} \; \mathsf{P}(\boldsymbol{f} \mid \boldsymbol{e}) \mathsf{P}(\boldsymbol{e})$$

  Translation Model
  Language Model

- Word-Based Translation
  Brown *et al.* (1988,1993)

Statistical Machine Translation

- Noisy Channel Model
  $\hat{\boldsymbol{e}} = \underset{\boldsymbol{e}}{\operatorname{argmax}} \mathsf{P}(\boldsymbol{f} \mid \boldsymbol{e}) \mathsf{P}(\boldsymbol{e})$

  Translation Model
  Language Model
  ...

- Phrase-Based Translation
  Och *et al.* (1999)
  Koehn *et al.* (2003)

Statistical Machine Translation

- Noisy Channel Model
  $$\hat{\boldsymbol{e}} = \underset{\boldsymbol{e}}{\operatorname{argmax}} \, \mathsf{P}(\boldsymbol{f} \,|\, \boldsymbol{e}) \mathsf{P}(\boldsymbol{e})$$

  Translation Model
  Language Model

- Phrase-Based Translation
  Och *et al.* (1999)
  Koehn *et al.* (2003)

Translation Model — $\mathrm{P}(\boldsymbol{f} \mid \boldsymbol{e})$

der Präsident $\rightarrow$ the president

Translation Model — P(f | e)

der Präsident → the president

 → 

Translation Model — P($\boldsymbol{f} \mid \boldsymbol{e}$)

der Präsident $\rightarrow$ the present

 $\rightarrow$ 

 $\rightarrow$ 

Translation Model — $\mathrm{P}(\boldsymbol{f} \mid \boldsymbol{e})$

der Präsident $\rightarrow$ the president

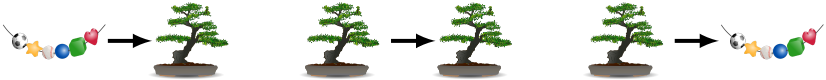
 

 $\rightarrow$ 

 $\rightarrow$ 

 $\rightarrow$ 

Translation Model — $P(f \mid e)$

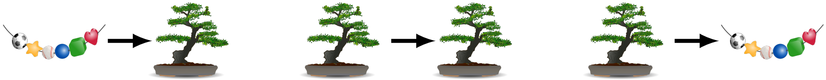


Translation Model — $P(f \mid e)$

Statistics + Syntactic Rules in the Translation Model

Abeillé *et al.*, 1990; Poutsma, 1998; Poutsma, 2000; Yamada & Knight, 2001; Yamada & Knight, 2002; Eisner, 2003; Gildea, 2003; Hearne & Way, 2003; Poutsma, 2003; Imamura *et al.*, 2004; Galley *et al.*, 2004; Graehl & Knight, 2004; Melamed, 2004; Ding & Palmer, 2005; Hearne, 2005; Quirk *et al.*, 2005; Cowan *et al.*, 2006; Galley *et al.*, 2006; Huang *et al.*, 2006; Liu *et al.*, 2006; Marcu *et al.*, 2006; Zollmann & Venugopal, 2006; Bod, 2007; DeNeefe *et al.*, 2007; Liu *et al.*, 2007; Chiang *et al.*, 2008; Lavie *et al.*, 2008; Mi & Huang, 2008; Mi *et al.*, 2008; Resnik, 2008; Shen *et al.*, 2008; Zhou *et al.*, 2008; Chiang, 2009; Hanneman & Lavie, 2009; Liu *et al.*, 2009; Chiang, 2010; Huang & Mi, 2010;

. . .

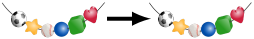


Language Model — P(e)

- Phrase-based Machine Translation
  - Linguistically naive
  - Most commonly-used statistical machine translation method
  - Outperforms syntactic TM systems for many language pairs

Statistics + Syntactic Rules in the Language Model

- Novel contribution of this work:
  - Technique for using any generative incremental parser as a syntactic language model
  - Incorporate our incremental syntactic language model into phrase-based machine translation



Language Model — P(e)

Estimate n-gram Language Model

$$\mathsf{P}(e_n \mid e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

Language Model — P(e)

Estimate n-gram Language Model

$$\mathsf{P}(e_n \mid e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In

Estimate n-gram Language Model

$$\mathsf{P}(e_n \mid e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \mid e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words

Language Model — P(e)

Estimate n-gram Language Model

$$\mathsf{P}(e_n \,|\, e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words ,

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \,|\, e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \mid e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \,|\, e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language

Language Model — $\mathsf{P}(\boldsymbol{e})$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \,|\, e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \,|\, e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model tries

Language Model — P(e)

Estimate n-gram Language Model

$$\mathsf{P}(e_n \,|\, e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model tries to

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \mid e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model tries to predict

Language Model — P(e)

Estimate n-gram Language Model

$$\mathsf{P}(e_n \mid e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model tries to predict the

Language Model — P(e)

Estimate n-gram Language Model

$$\mathsf{P}(e_n \,|\, e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model tries to predict the next

Language Model — P(e)

Estimate n-gram Language Model

$$\mathsf{P}(e_n \mid e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model tries to predict the next word

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \mid e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model tries to predict the next word in

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \mid e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model tries to predict the next word in a

Language Model — $\mathrm{P}(e)$

Estimate n-gram Language Model

$$\mathrm{P}(e_n \,|\, e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model tries to predict the next word in a sequence

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \,|\, e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model tries to predict the next word in a sequence of

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \mid e_1 \ldots e_{n-1}) = \frac{C(e_1 ... e_n)}{C(e_1 ... e_{n-1})}$$

In other words , an n-gram language model tries to predict the next word in a sequence of words

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \,|\, e_1 \ldots e_{n-1}) = \frac{C(e_1...e_n)}{C(e_1...e_{n-1})}$$

In other words , an n-gram language model tries to predict the next word in a sequence of words .

Language Model — $\mathsf{P}(e)$

Estimate n-gram Language Model

$$\mathsf{P}(e_n \,|\, e_1 \ldots e_{n-1}) = \frac{C(e_1...e_n)}{C(e_1...e_{n-1})}$$

In other words , an n-gram language model tries to predict the next word in a sequence of words .

- Widely used in speech recognition & machine translation
- Can be trained on a corpus of monolingual data
- Variety of backoff and smoothing techniques to account for words not encountered during training

The

<s>

Language Model — P(e)

The pictures

<s> The

Language Model — P(e)

The pictures of

The pictures

Language Model — $P(e)$

The pictures of the

pictures of

Language Model — $\mathrm{P}(e)$

The pictures of the old

of the

Language Model — P(e)

The pictures of the old man

the old

Language Model — P(e)

The pictures of the old man is

old man

Language Model — P(e)

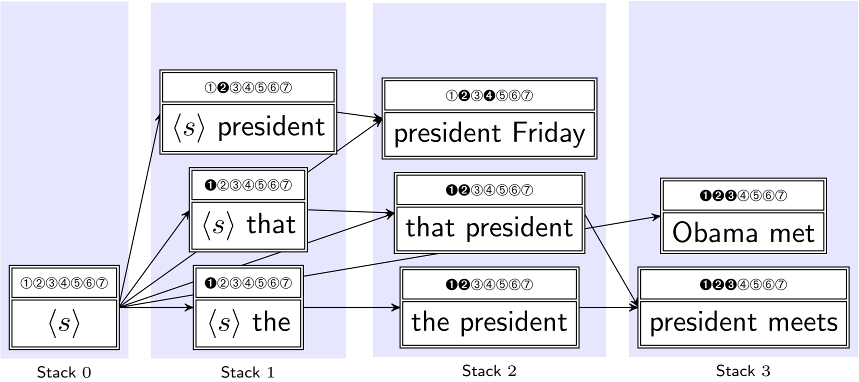
The pictures of the old man ~~is~~ are

old man

Phrase-Based Translation

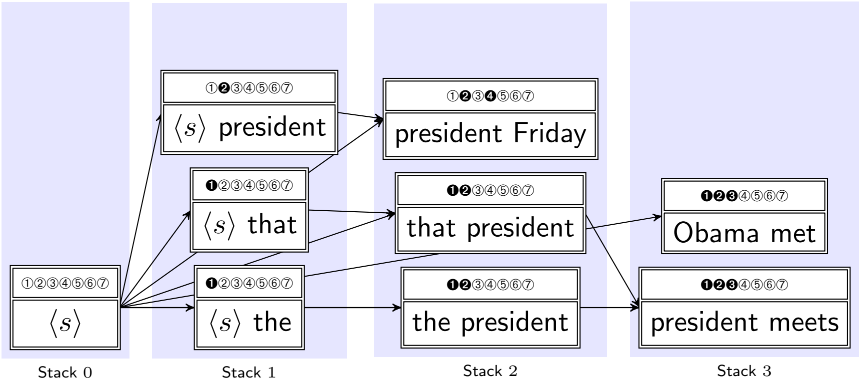
Der Präsident trifft am Freitag den Vorstand

The president meets the board on Friday



Definition

$\tilde{\boldsymbol{\tau}}_{t_h}$ represents parses of the partial translation at node h in stack t



Parsing

The president meets the board on Friday.

Parsing

The president meets the board on Friday.

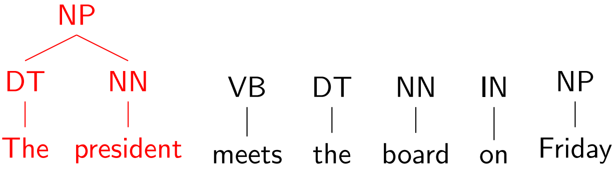
Parsing

Bottom-up parsing requires entire sentence

DT	NN	VB	DT	NN	IN	NP
The	president	meets	the	board	on	Friday

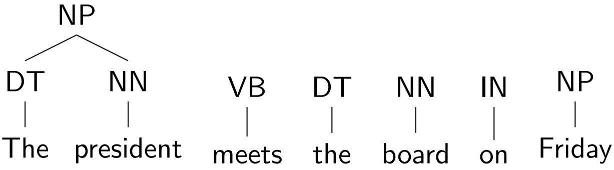
The president meets the board on Friday

Parsing



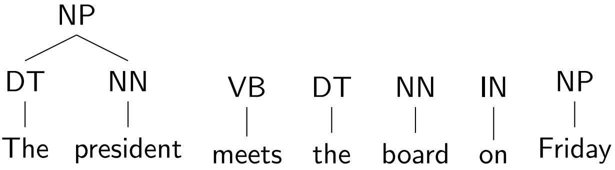
The president meets the board on Friday

Parsing



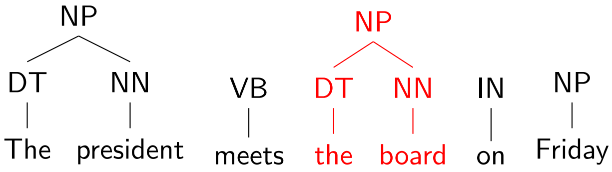
The president meets the board on Friday

Parsing



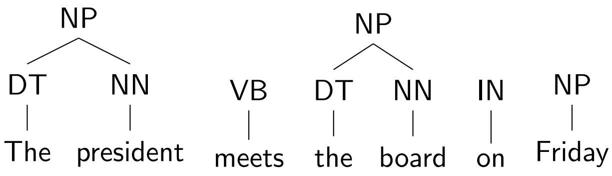
The president meets the board on Friday

Parsing



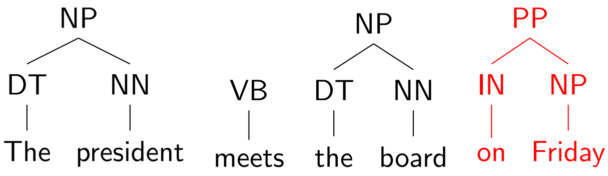
The president meets the board on Friday

Parsing



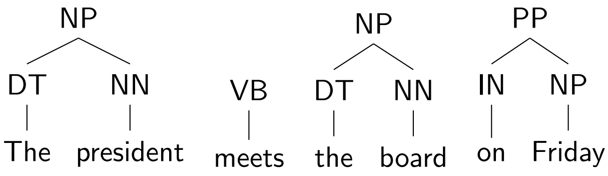
The president meets the board on Friday

Parsing



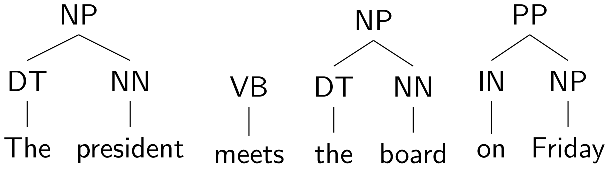
The president meets the board on Friday

Parsing



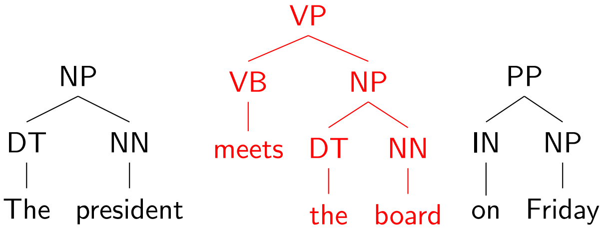
The president meets the board on Friday

Parsing



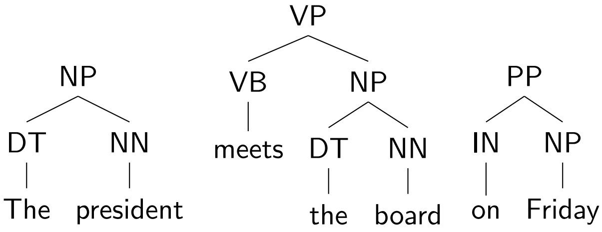
The president meets the board on Friday

Parsing



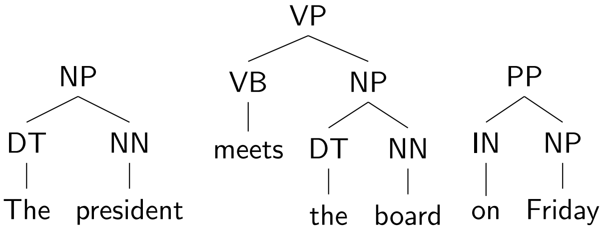
The president meets the board on Friday

Parsing



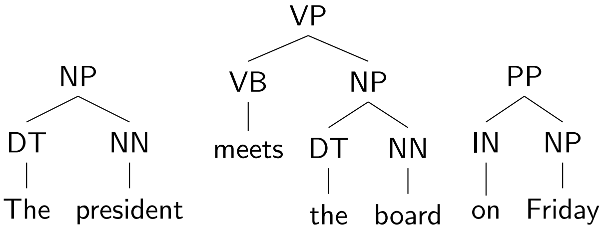
The president meets the board on Friday

Parsing



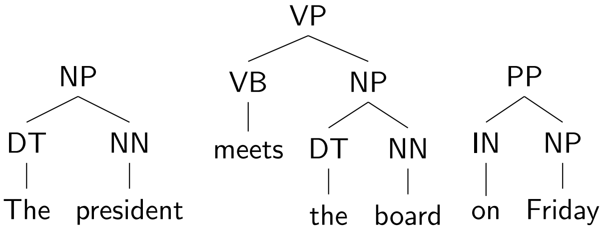
The president meets the board on Friday

Parsing



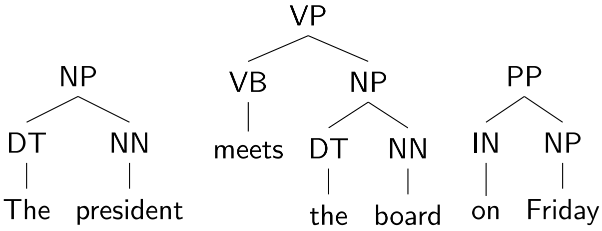
The president meets the board on Friday

Parsing



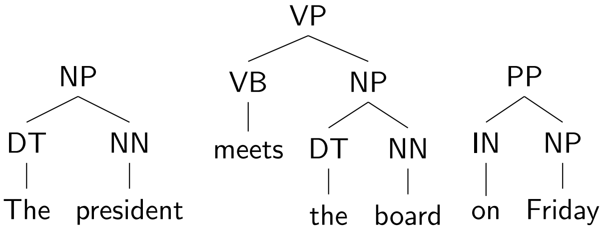
The president meets the board on Friday

Parsing



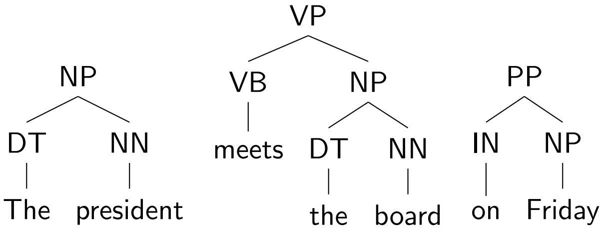
The president meets the board on Friday

Parsing

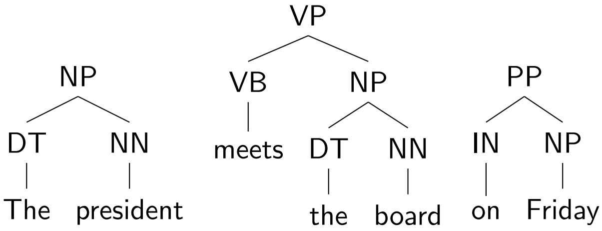


The president meets the board on Friday

Parsing

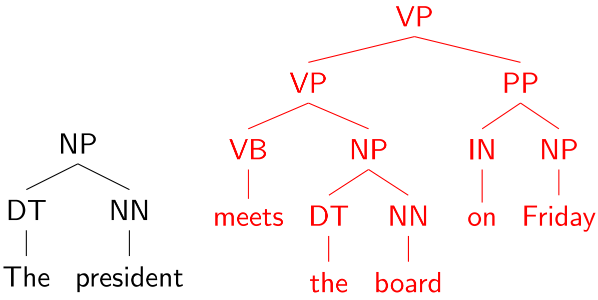


The president meets the board on Friday



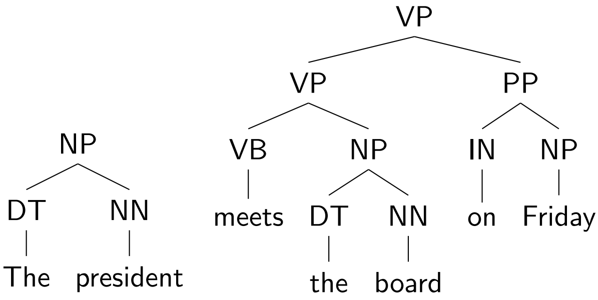
The president meets the board on Friday

Parsing



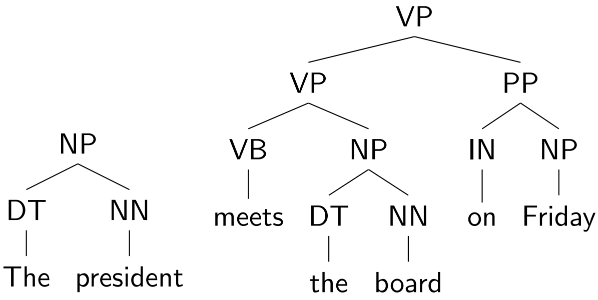
The president meets the board on Friday

Parsing



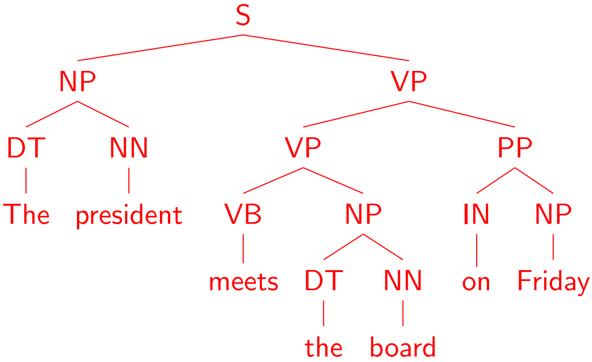
The president meets the board on Friday

Parsing



The president meets the board on Friday

Parsing



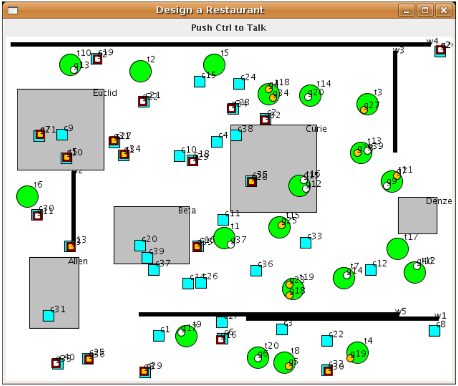
The president meets the board on Friday

Incremental Parsing

- Humans hear language incrementally
- Humans process language incrementally
- Incremental parsers have nice pyscholinguistic properties
- Incremental parsers can process partial sentences

Incremental Parsing

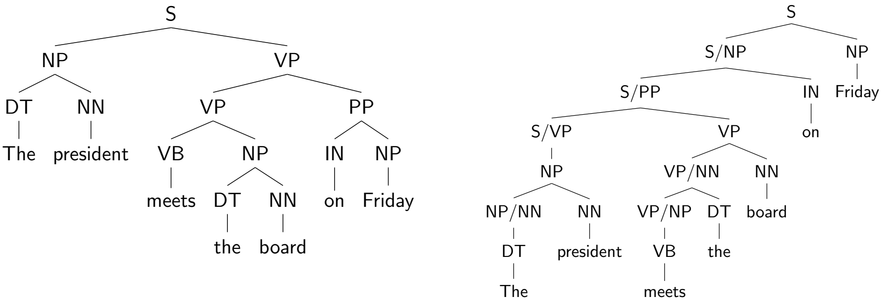
- Spoken language interfaces (Schwartz et al, 2009)



- Handling realistic disfluent spoken input (Miller et al, 2009)
- Modelling reading time (Wu et al, 2010)
- Coreference resolution (ongoing)

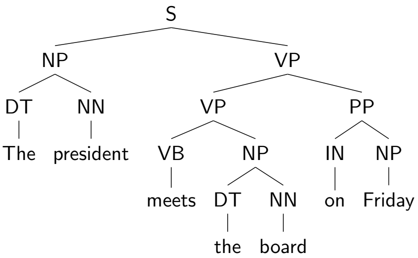
Right-corner Incremental Parsing

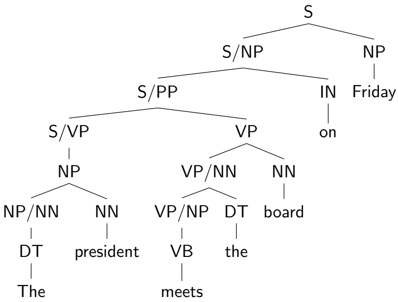
Transform right-expanding sequences of constituents



Right-corner Incremental Parsing

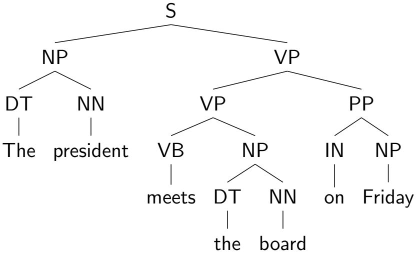
Transform right-expanding sequences of constituents
into left-expanding sequences of incomplete constituents

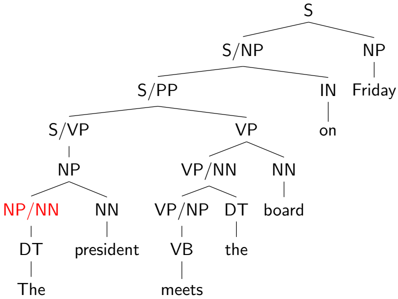




Right-corner Incremental Parsing

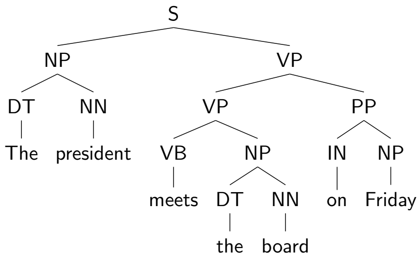
Transform right-expanding sequences of constituents
into left-expanding sequences of incomplete constituents

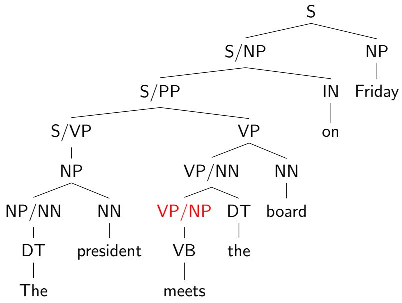




Right-corner Incremental Parsing

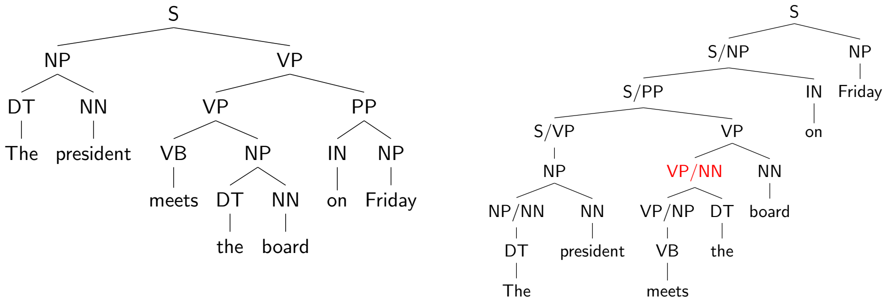
Transform right-expanding sequences of constituents
into left-expanding sequences of incomplete constituents





Right-corner Incremental Parsing

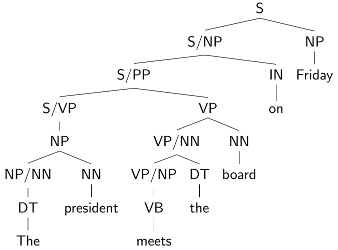
Transform right-expanding sequences of constituents
into left-expanding sequences of incomplete constituents

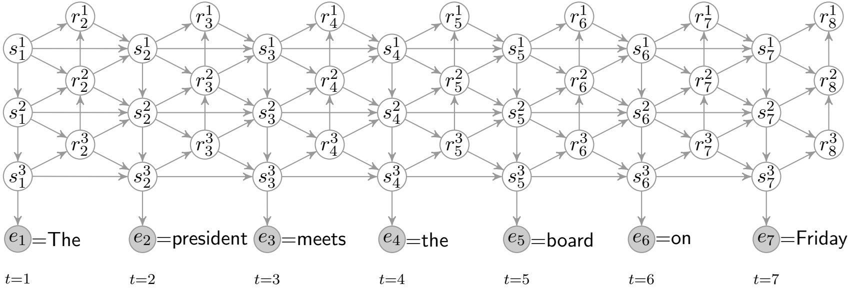


Right-corner Incremental Parsing using HHMM

Hierarchical Hidden Markov Model

- Circles denote
  hidden random variables
- Edges denote
  conditional dependencies
- Shaded circles denote
  observed values

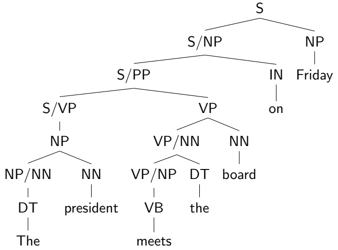


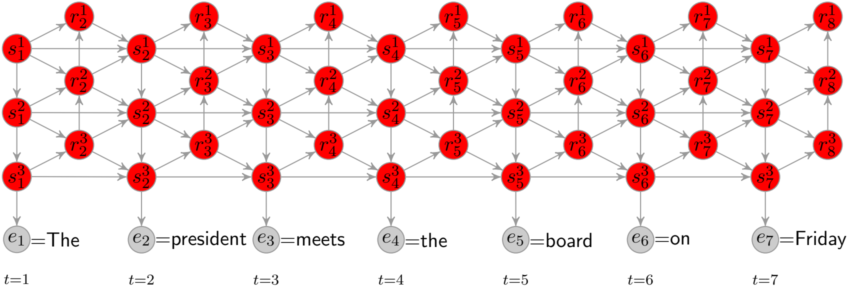


Right-corner Incremental Parsing using HHMM

Hierarchical Hidden Markov Model

- Circles denote
  hidden random variables
- Edges denote
  conditional dependencies
- Shaded circles denote
  observed values

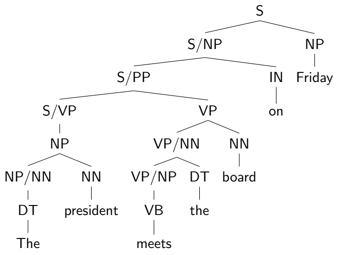


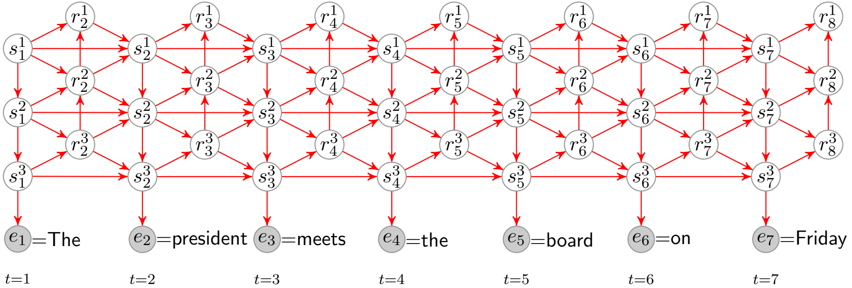


Right-corner Incremental Parsing using HHMM

Hierarchical Hidden Markov Model

- Circles denote
  hidden random variables
- Edges denote
  conditional dependencies
- Shaded circles denote
  observed values

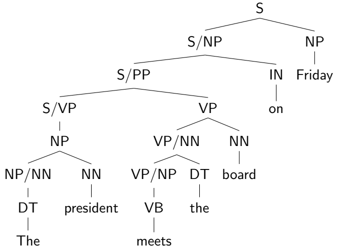


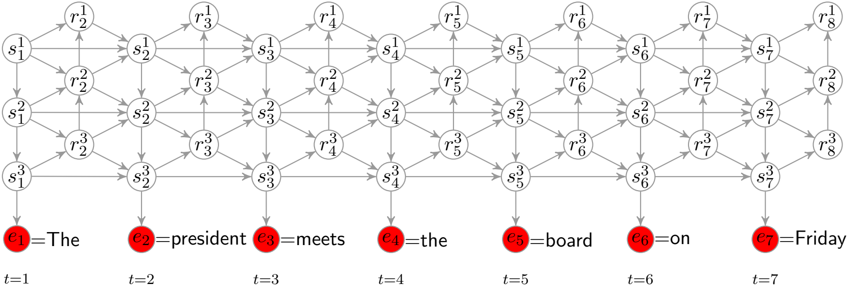


Right-corner Incremental Parsing using HHMM

Hierarchical Hidden Markov Model

- Circles denote
  hidden random variables
- Edges denote
  conditional dependencies
- Shaded circles denote
  observed values



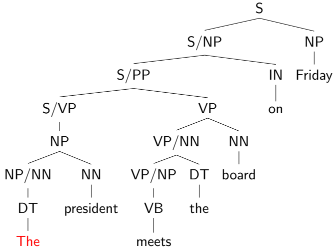


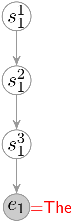
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



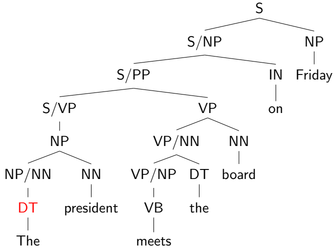


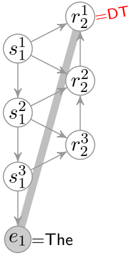
Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental” CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



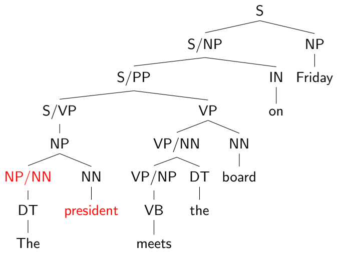


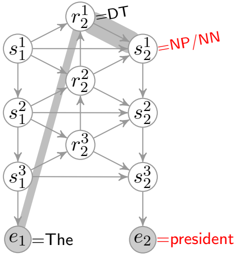
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



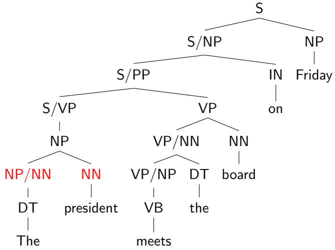


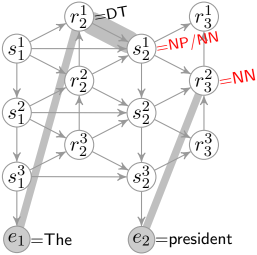
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



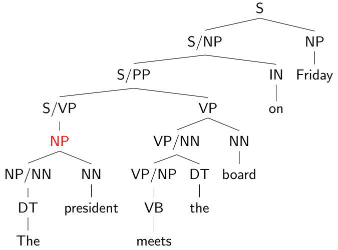


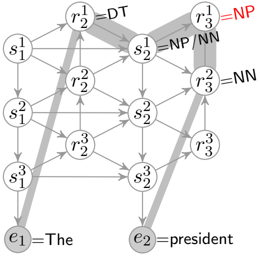
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



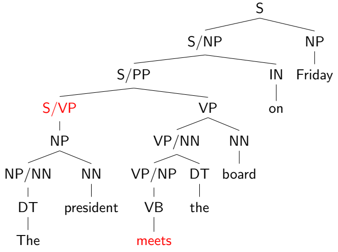


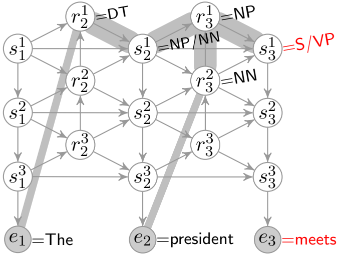
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



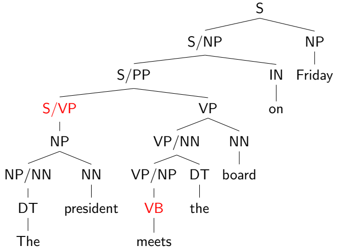


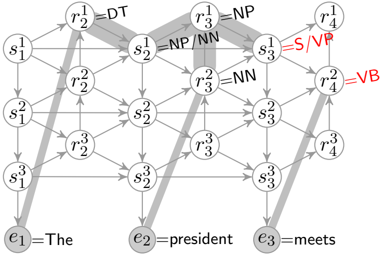
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



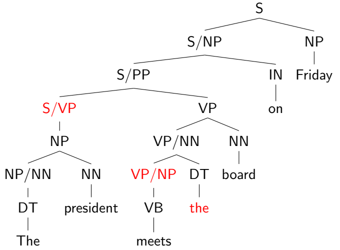


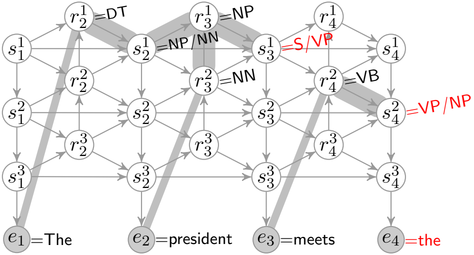
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



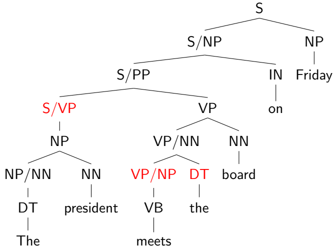


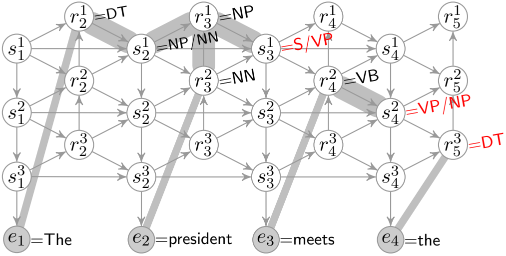
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



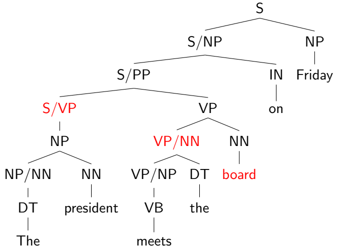


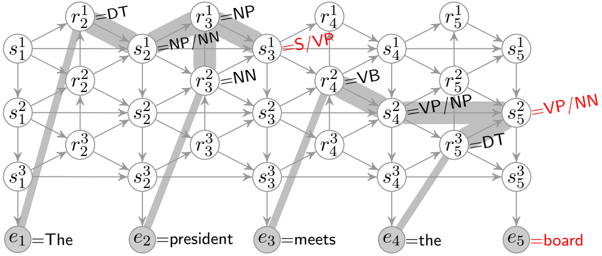
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



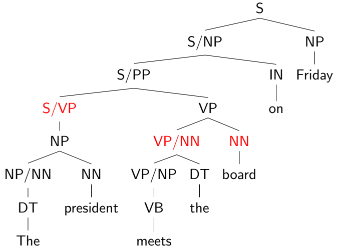


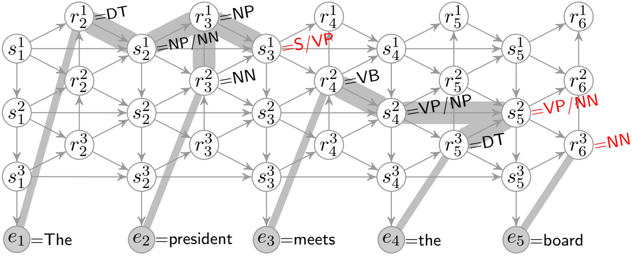
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



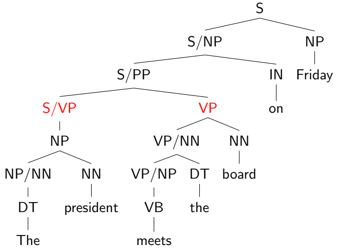


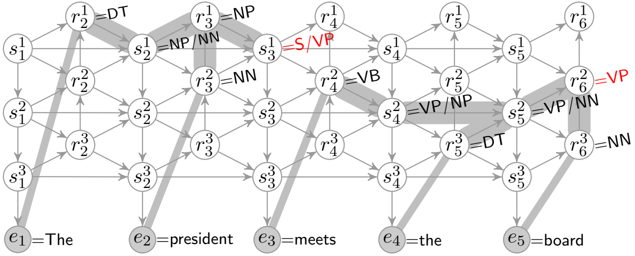
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental"
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree → Path



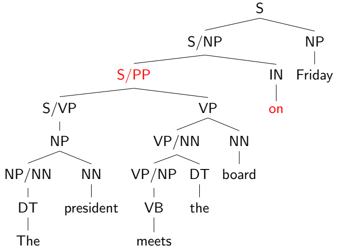


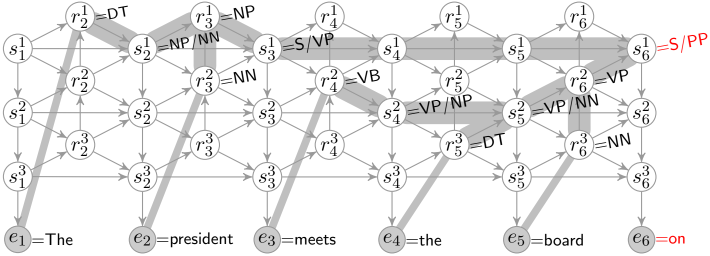
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



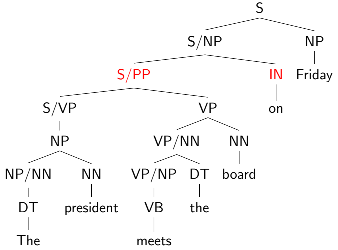


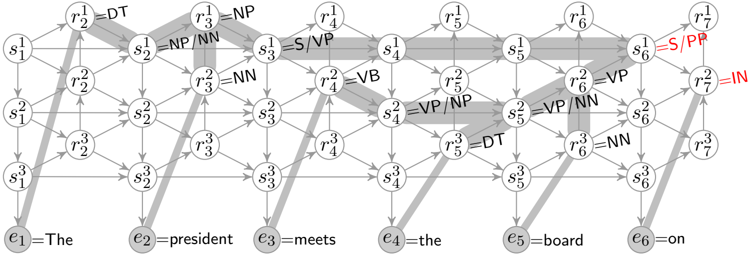
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



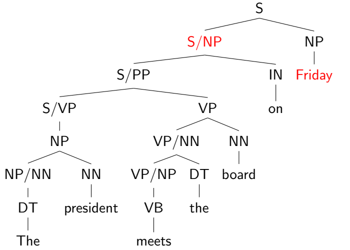


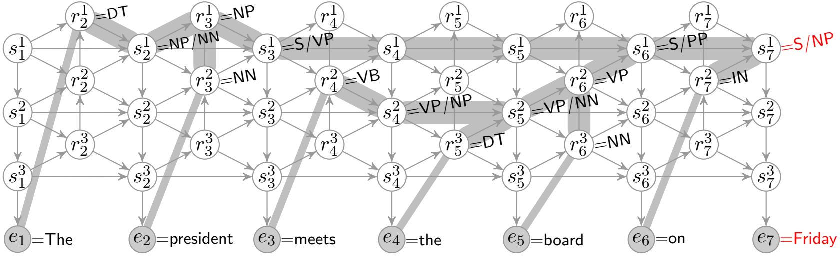
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



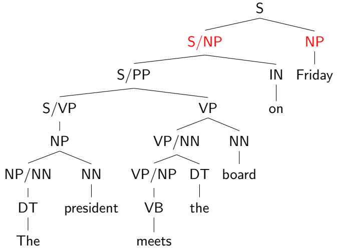


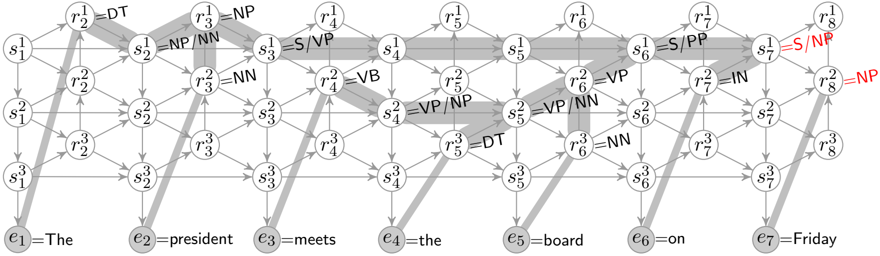
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

Isomorphic Tree → Path



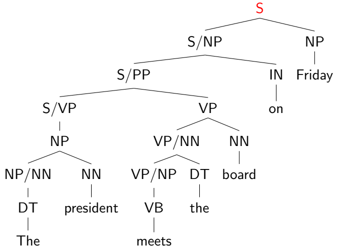


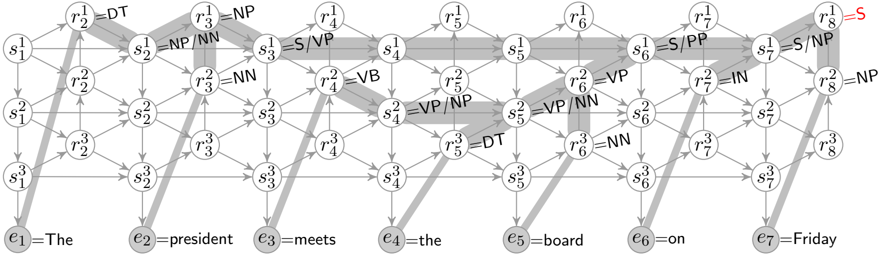
Right-corner Incremental Parsing using HHMM

Analogous to "Maximally Incremental" CCG Parsing

Analogous to Probabilistic Push-Down Automata

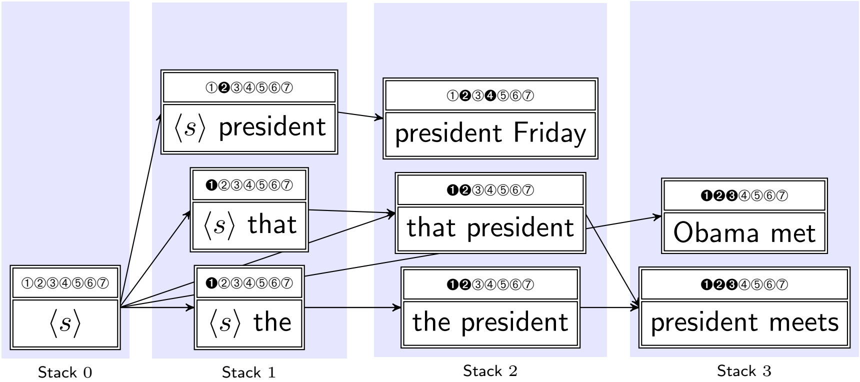
Isomorphic Tree → Path





Phrase-Based Translation is also Incremental

Der Präsident trifft am Freitag den Vorstand
The president meets the board on Friday



Phrase-Based Translation is also Incremental

Der Präsident trifft am Freitag den Vorstand

①②③④⑤⑥⑦

$\langle s \rangle$

Stack 0

Stack 1

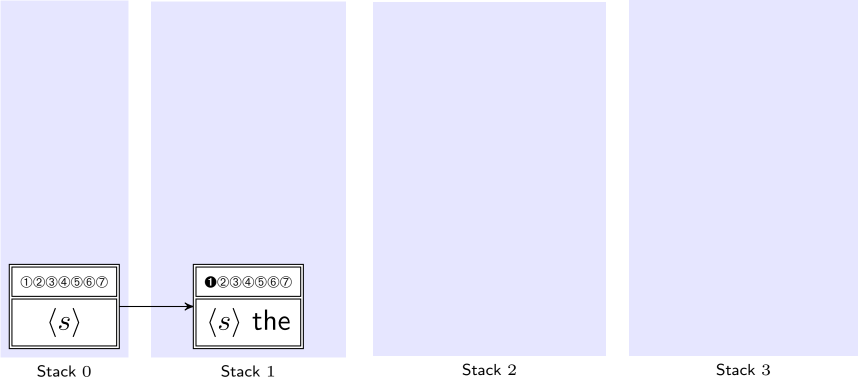
Stack 2

Stack 3

Phrase-Based Translation is also Incremental

Der Präsident trifft am Freitag den Vorstand

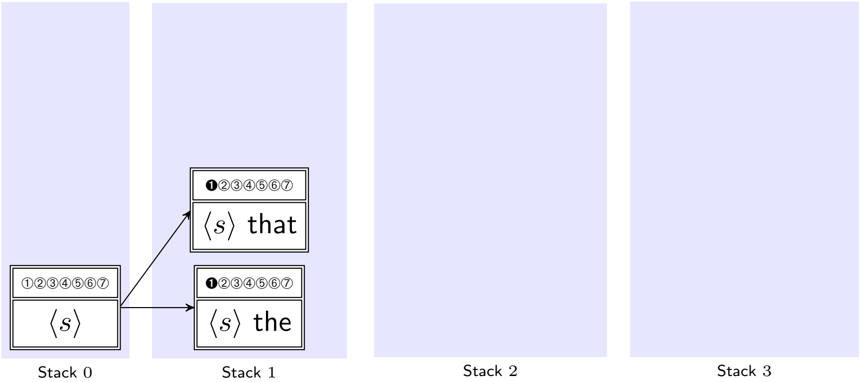
The



Phrase-Based Translation is also Incremental

Der Präsident trifft am Freitag den Vorstand

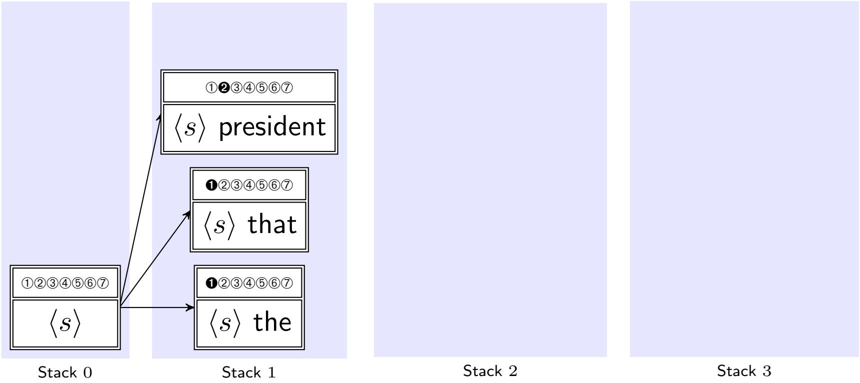
That



Phrase-Based Translation is also Incremental

Der *Präsident* *trifft am Freitag den Vorstand*

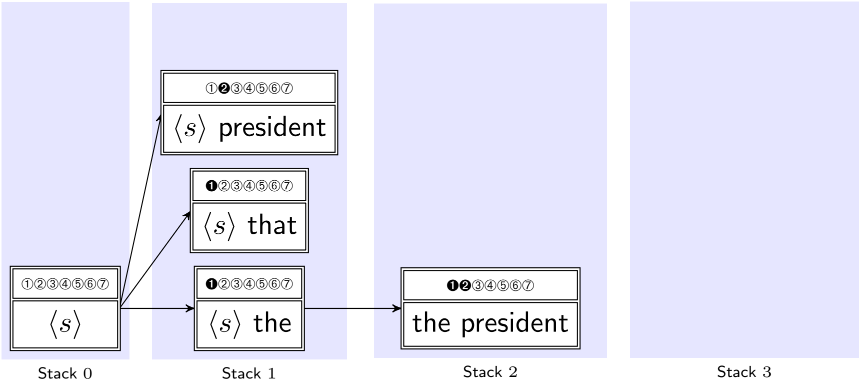
President



Phrase-Based Translation is also Incremental

Der Präsident *trifft am Freitag den Vorstand*

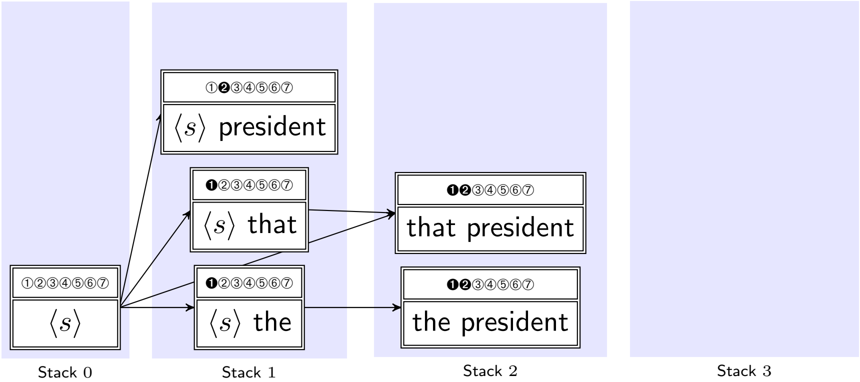
The president



Phrase-Based Translation is also Incremental

Der Präsident trifft am Freitag den Vorstand

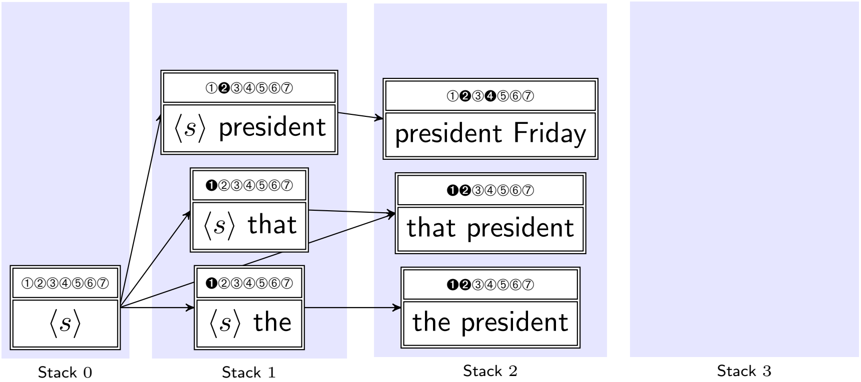
That president



Phrase-Based Translation is also Incremental

Der Präsident trifft am Freitag den Vorstand

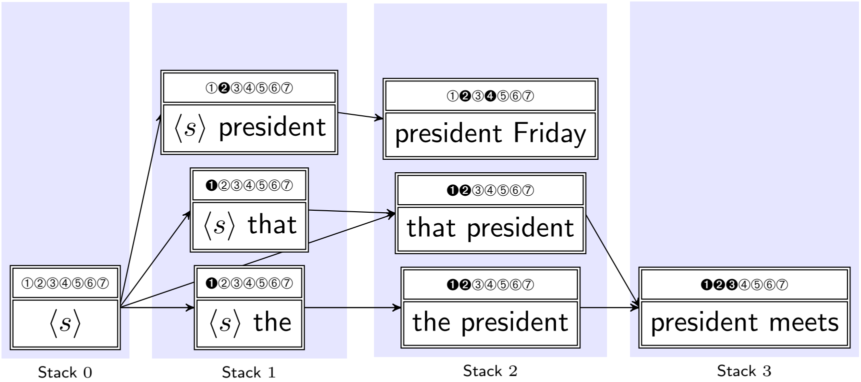
President Friday



Phrase-Based Translation is also Incremental

Der Präsident trifft am Freitag den Vorstand

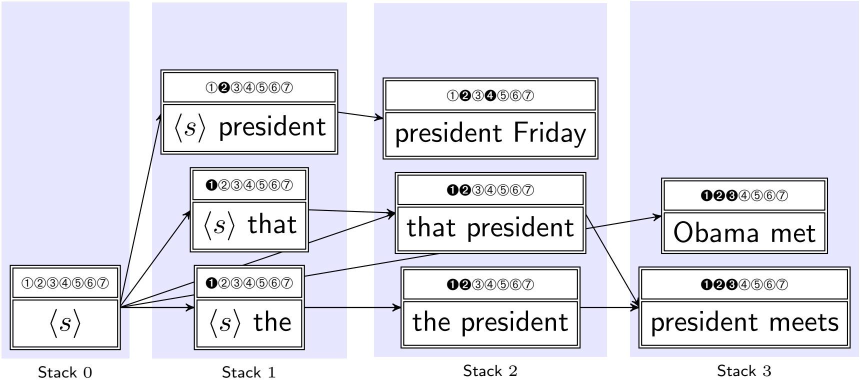
The president meets



Phrase-Based Translation is also Incremental

Der *Präsident trifft* *am Freitag den Vorstand*

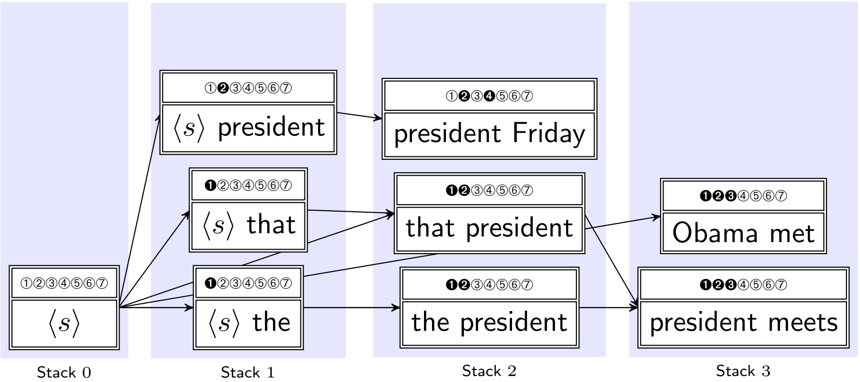
Obama met



Phrase-Based Translation is also Incremental

Definition

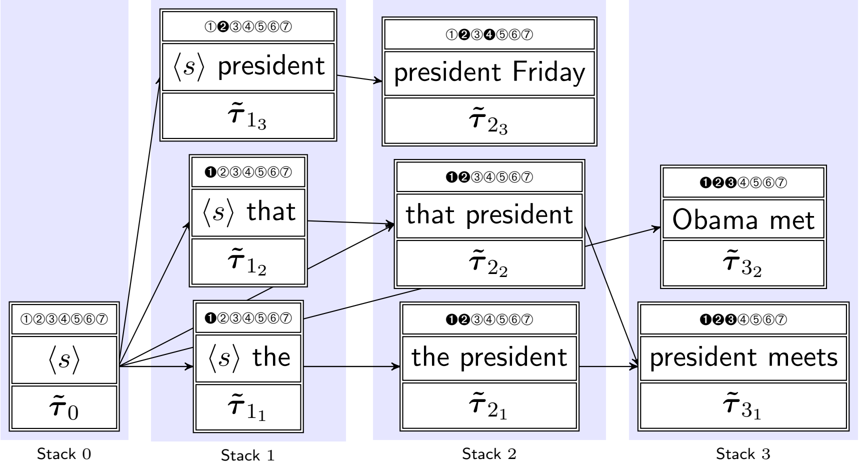
$\tilde{\boldsymbol{\tau}}_{t_h}$ represents parses of the partial translation at node h in stack t



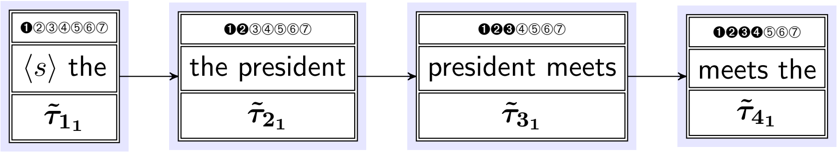
Phrase-Based Translation with Syntactic LM

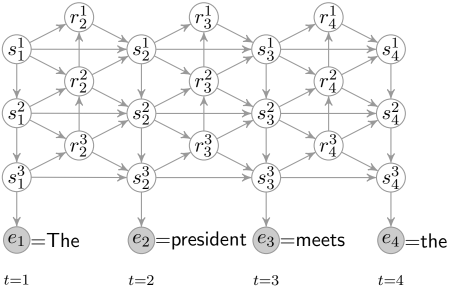
Definition

$\tilde{\boldsymbol{\tau}}_{t_h}$ represents parses of the partial translation at node h in stack t

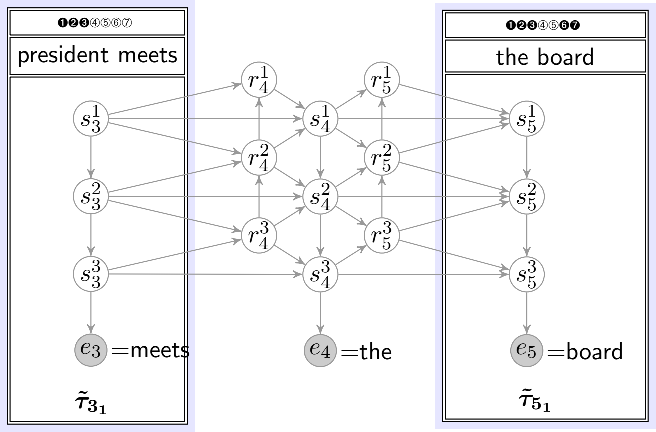


Integrate Parser into Phrase-based Decoder

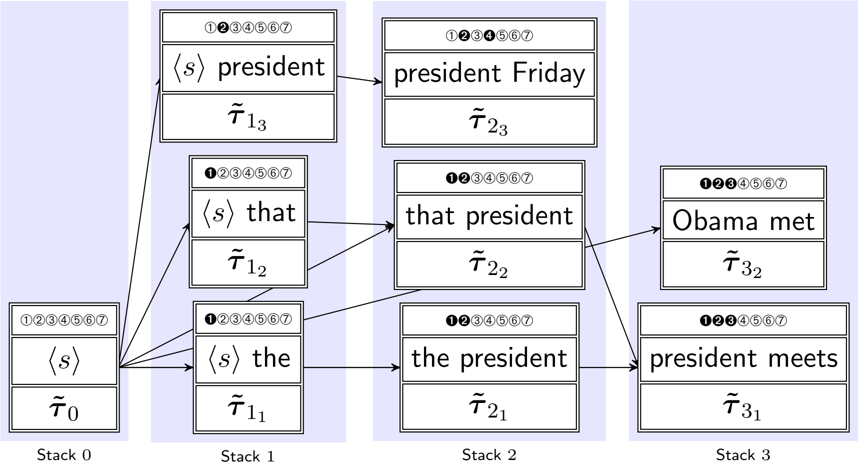




Integrate Parser into Phrase-based Decoder



Syntactic Language Model Guides Pruning



Syntactic Language Model Guides Pruning

①❷③❹⑤⑥⑦

president Friday

$\tilde{\boldsymbol{\tau}}_{2_3}$

❶❷③④⑤⑥⑦

that president

$\tilde{\boldsymbol{\tau}}_{2_2}$

❶❷③④⑤⑥⑦

the president

$\tilde{\boldsymbol{\tau}}_{2_1}$

Stack 2

Syntactic Language Model Guides Pruning

①❷③❹⑤⑥⑦

president Friday

$\tilde{\boldsymbol{\tau}}_{2_3}$

❶❷③④⑤⑥⑦

that president

$\tilde{\boldsymbol{\tau}}_{2_2}$

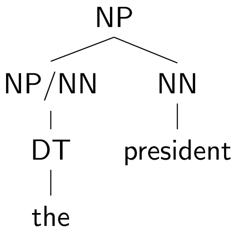
❶❷③④⑤⑥⑦

the president

$\tilde{\boldsymbol{\tau}}_{2_1}$

Stack 2

$\tilde{\boldsymbol{\tau}}_{2_1}$



$$\mathrm{P}(\tilde{\boldsymbol{\tau}}_{2_1}) = 0.15$$

Syntactic Language Model Guides Pruning

①❷③❹⑤⑥⑦

president Friday

$\tilde{\boldsymbol{\tau}}_{2_3}$

❶❷③④⑤⑥⑦

that president

$\tilde{\boldsymbol{\tau}}_{2_2}$

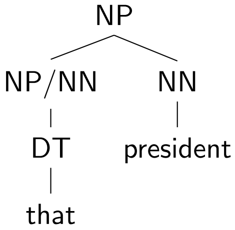
❶❷③④⑤⑥⑦

the president

$\tilde{\boldsymbol{\tau}}_{2_1}$

Stack 2

$\tilde{\boldsymbol{\tau}}_{2_2}$



$\mathsf{P}(\tilde{\boldsymbol{\tau}}_{2_2}) = 0.12$

Syntactic Language Model Guides Pruning

①❷③❹⑤⑥⑦
president Friday
$\tilde{\boldsymbol{\tau}}_{2_3}$

❶❷③④⑤⑥⑦
that president
$\tilde{\boldsymbol{\tau}}_{2_2}$

❶❷③④⑤⑥⑦
the president
$\tilde{\boldsymbol{\tau}}_{2_1}$

Stack 2

$\tilde{\boldsymbol{\tau}}_{2_3}$

NN NN

president Friday

$$\mathrm{P}(\tilde{\boldsymbol{\tau}}_{2_3}) = 0.05$$

Syntactic Language Model Guides Pruning

Our work presents a novel mechanism for incorporating syntax into the language model of phrase-based machine translation

Evaluation

How do we know if the syntactic language model is good?

Evaluation

How do we know if the syntactic language model is good?

- BLEU
- Perplexity
- Manual

Evaluation — BLEU

Experiment

- NIST OpenMT 2008 Urdu-English data set
- Moses with standard phrase-based translation model
- Tuning and testing restricted to sentences $\leq$ 40 words long
- Results reported on devtest set
- n-gram LM is WSJ 5-gram LM

Evaluation — BLEU

BLEU

- Modified precision metric for assessing translation quality
- Measures n-gram matches against reference translations
- Higher BLEU scores are better
- Does **not** measure syntactic well-formedness

Evaluation — BLEU

BLEU

- Modified precision metric for assessing translation quality
- Measures n-gram matches against reference translations
- Higher BLEU scores are better
- Does **not** measure syntactic well-formedness

Moses LM(s)	reordering limit=10	reordering limit=20
n-gram only	21.67	21.88
HHMM + n-gram	21.44	21.93

Evaluation — Perplexity

Perplexity

- Standard measure of language model quality
- Reports how surprised a model is by test data
- Lower perplexity is better
- Calculated using log base b for a test set of T tokens.

$$ppl = b^{\frac{-log_b \mathsf{P}(e_1 \ldots e_T)}{T}}$$

Evaluation — Perplexity

Language models trained on WSJ Treebank corpus

LM	In-domain Perplexity	Out-of-domain Perplexity
WSJ 5-gram LM	**232**	1262
WSJ Syntactic LM	385	**529**

Evaluation — Perplexity

Language models trained on WSJ Treebank corpus

LM	In-domain Perplexity	Out-of-domain Perplexity
WSJ 5-gram LM	**232**	1262
WSJ Syntactic LM	385	**529**
Interpolated WSJ 5-gram + WSJ SynLM	**209**	**225**

Evaluation — Perplexity

Language models trained on WSJ Treebank corpus
...and n-gram model for larger English Gigaword corpus.

LM	In-domain Perplexity	Out-of-domain Perplexity
WSJ 5-gram LM	**232**	1262
WSJ Syntactic LM	385	529
Interpolated WSJ 5-gram + WSJ SynLM	<u>**209**</u>	225
Gigaword 5-gram	258	**312**
Interpolated Gigaword 5-gram + WSJ SynLM	222	<u>**123**</u>

Evaluation — Manual

Manual Examination

- Actually look at the translations
- Gold standard for measuring quality
- Assess syntactic well-formedness

Evaluation — Manual

ID	Segment 624, Document "devtest" [Δ_{BLEU}=-0.15]
Source	پر ' وقت لکھے گا تاریخ کا فیصلہ ' .
Reference (reference0)	but ' time will recount the judgment of history ' .
Reference (reference1)	but ' time will write the judgment of history ' .
Reference (reference2)	but time will decide what history will write in the end .
Reference (reference3)	but time will write the decision of the history .
Hypothesis (ngram)	the decision of history **written on ' time will '** . [0.29]
Hypothesis (hhmm)	**' time will write on** the decision of history . [0.44]

Evaluation — Manual

ID	Segment 744, Document "devtest" [Δ_{BLEU}=-0.23]
Source	ملاقات میں حرج نہیں .
Reference (reference0)	there is nothing wrong in meeting .
Reference (reference1)	there is no problem in meeting .
Reference (reference2)	there is no harm in meeting with him .
Reference (reference3)	there are no problems with this meeting .
Hypothesis (ngram)	in the meeting **, is not** . [0.09]
Hypothesis (hhmm)	**no harm** in the meeting . [0.32]

Evaluation — Manual

ID	Segment 561, Document "devtest" [Δ_{BLEU}=-0.21]
Source	ہر ا نسان کو معاشرے میں ا پنی ذمے داری سمجھنا چاہئے
Reference (reference0)	everyone must recognize his responsibility in the society
Reference (reference1)	every person should realize ones responsibility in the society .
Reference (reference2)	everyone in society should do his duty .
Reference (reference3)	every man should understand his responsibilities to society .
Hypothesis (ngram)	**the society should understand their in** every human being claimed responsibility [0.13]
Hypothesis (hhmm)	every human being claimed responsibility **in the society should understand** [0.34]

Conclusion

- Many others have incorporated syntax into translation model

Conclusion

- Many others have incorporated syntax into translation model
- Phrase-based machine translation uses a syntactically naive translation model

Conclusion

- Many others have incorporated syntax into translation model
- Phrase-based machine translation uses a syntactically naive translation model
- Our work presents a novel mechanism for incorporating syntax into the language model

Conclusion

- Many others have incorporated syntax into translation model
- Phrase-based machine translation uses a syntactically naive translation model
- Our work presents a novel mechanism for incorporating syntax into the language model
- Use any generative incremental parser as syntactic language model

Conclusion

- Many others have incorporated syntax into translation model
- Phrase-based machine translation uses a syntactically naive translation model
- Our work presents a novel mechanism for incorporating syntax into the language model
- Use any generative incremental parser as syntactic language model
- Straightforward and natural mechanism for integrating syntax into phrase-based machine translation

Thanks

Thank you!

Thanks to . . .

- My wife & our children
- My incredibly supportive family and friends
- My advisor William Schuler & the UMN NLP lab
- Chris Callison-Burch & the Johns Hopkins CLSP & COE
- Georganne Tolaas, without whom nothing is possible
- My colleagues in the Speech & Language lab at AFRL
- Chris, Philipp, Josh, Chris, Adam, Phil, Hieu, Barry, Jon, Ondřej, Omar, Ken & the rest of the MT Marathon hackers
- My former colleagues at IBM Rochester
- Ann Copestake, my Cambridge M.Phil classmates, & the Gates Cambridge Foundation
- My committee

Acknowledgments

- Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Air Force.

Incremental Parser as Syntactic LM Feature

$$\hat{\boldsymbol{e}} = \underset{\boldsymbol{e}}{\operatorname{argmax}} \ \exp \sum_j \lambda_j h_j(\boldsymbol{e}, \boldsymbol{f})$$

$$\boldsymbol{\lambda} = \text{Set of } j \text{ feature weights}$$

$$\boldsymbol{h} = \begin{cases} \text{Phrase-based translation model} \\ n\text{-gram LM} \\ \text{Distortion model} \\ \vdots \\ \text{Syntactic LM } \mathsf{P}(\tilde{\boldsymbol{\tau}}_{t_h}) \end{cases}$$

Results — Manual Examination

ID	Segment 103, Document "devtest" [Δ_{BLEU}=0.68]
Source	حکومت کے وعدے ? ? ?
Reference (reference0)	the promises of the government ? ? ?
Reference (reference1)	government 's promises ? ? ?
Reference (reference2)	the promises of the govt . ? ? ?
Reference (reference3)	government promises ? ? ?
Hypothesis (ngram)	the government ? **? ?** [1.00]
Hypothesis (hhmm)	the government ? **promise . .** [0.32]

Results — Manual Examination

ID	Segment 158, Document "devtest" [Δ_{BLEU}=-0.34]
Source	موجودہ چیف جسٹس کے خلاف دوبارہ ریفرنس دائر نہیں کیا جاسکتا ، سعیدالزمان صدیقی
Reference (reference0)	reference can not be filed again against the present chief justice , saeed uz zaman siddiqui
Reference (reference1)	another reference can not be filed against present chief justice : saeeduz zaman siddiqi
Reference (reference2)	saiduz zaman siddiqi : a second reference can not be filed against the present chief justice
Reference (reference3)	saiduz zaman siddiqi : a second reference can not be filed against the present chief justice
Hypothesis (ngram)	the chief justice **of سعیدالزمان siddiqui** can not **file a** reference again **.** [0.10]
Hypothesis (hhmm)	the chief justice can not **be filed against the** reference again **, سعیدالزمان siddiqui** [0.44]