

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

Lane Schwartz

Statistical Machine Translation

- Noisy Channel Model

$$\hat{e} = \operatorname{argmax}_e P(f | e)P(e)$$



Statistical Machine Translation

- Noisy Channel Model

$$\hat{e} = \operatorname{argmax}_e P(\mathbf{f} | \mathbf{e}) P(\mathbf{e})$$

Translation Model

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



Statistical Machine Translation

- Noisy Channel Model

$$\hat{e} = \operatorname{argmax}_e P(\mathbf{f} | \mathbf{e}) P(\mathbf{e})$$

Translation Model
Language Model

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



Statistical Machine Translation

- Maximum Entropy Model
$$\hat{e} = \operatorname{argmax}_e \exp \sum_j \lambda_j h_j(e, f)$$

Translation Model

Language Model

...

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



Translation Model — $P(f | e)$

der Präsident → the president



Translation Model — $P(f | e)$

der Präsident → the president



Translation Model — $P(f | e)$

der Präsident → the president



Translation Model — $P(f | e)$

der Präsident → the president



Translation Model — $P(f | e)$



Motivation

○○●○○○

Incremental Parsing

○○○○○○○

Integration

○○○○○○○○○

Results

○○○○○○○

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

Lane Schwartz

Translation Model — $P(f | e)$

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;

...

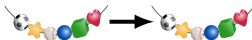


Language Model — $P(e)$

- 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



Estimate n -gram Language Model

$$P(e_n | e_1 \dots e_{n-1}) = \frac{C(e_1 \dots e_n)}{C(e_1 \dots e_{n-1})}$$

- 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

Language Model — $P(e)$

The

<S>

Language Model — $P(e)$

The pictures

<s> The

The pictures of

The pictures

The pictures of the

pictures of

Language Model — $P(e)$

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 — $P(e)$

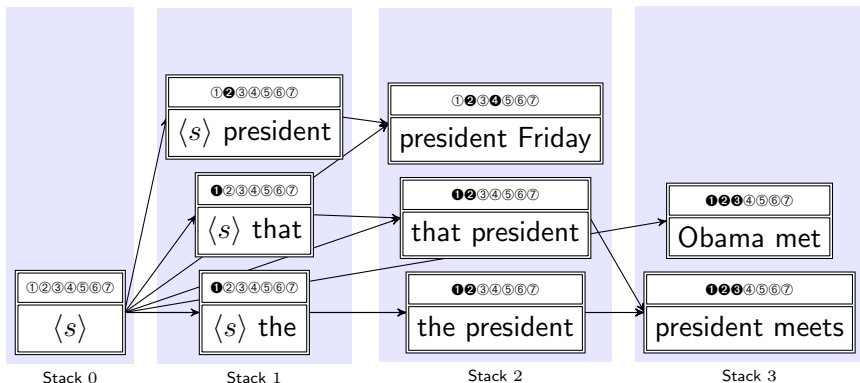
The pictures of the old man is are

old man

Phrase-Based Translation

Der Präsident trifft am Freitag den Vorstand

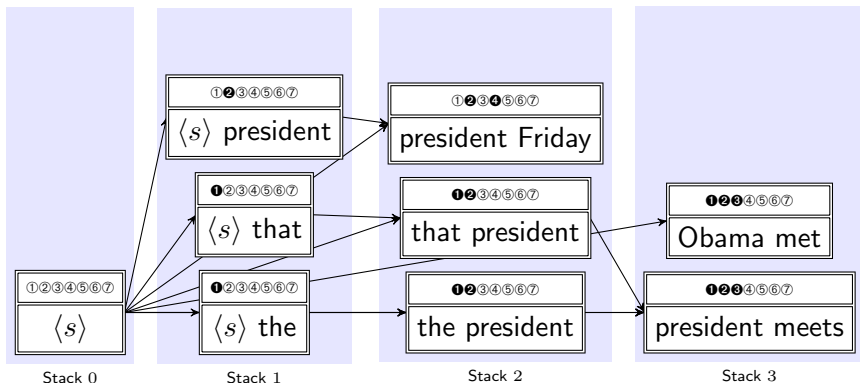
The president meets the board on Friday



Phrase-Based Translation

Definition

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



The president meets the board on Friday.





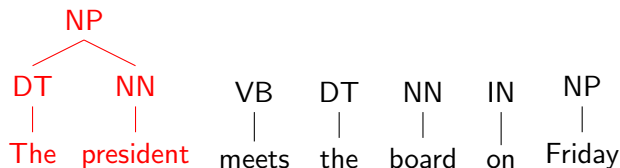
The president meets the board on Friday.

Bottom-up parsing requires **entire** sentence

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

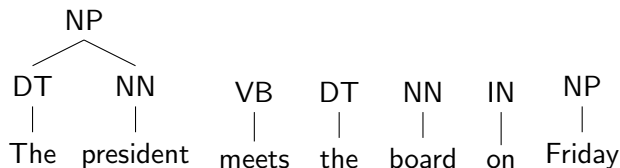
The president meets the board on Friday

Parsing



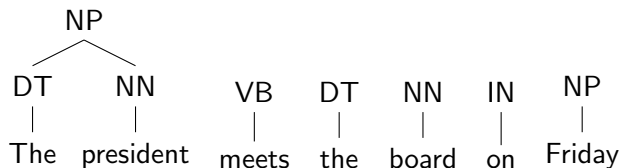
The president meets the board on Friday

Parsing



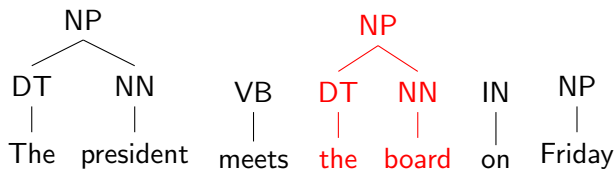
The **president meets** the board on Friday

Parsing



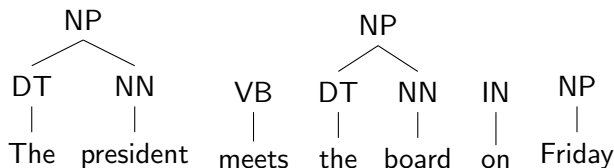
The president **meets the** board on Friday

Parsing



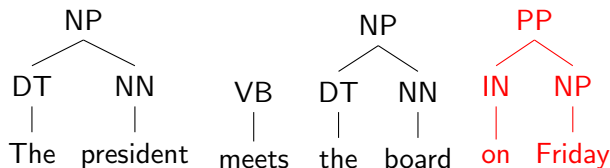
The president meets **the board** on Friday

Parsing



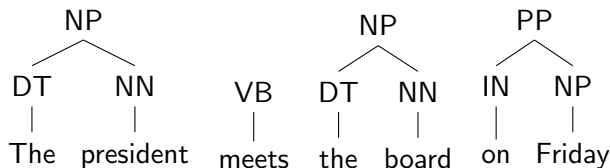
The president meets the **board on** Friday

Parsing



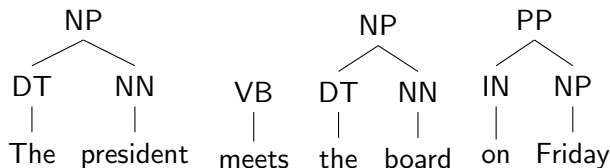
The president meets the board on Friday

Parsing



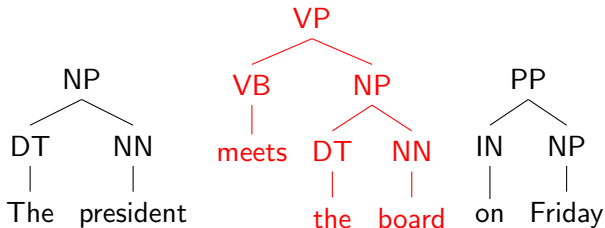
The president meets the board on Friday

Parsing



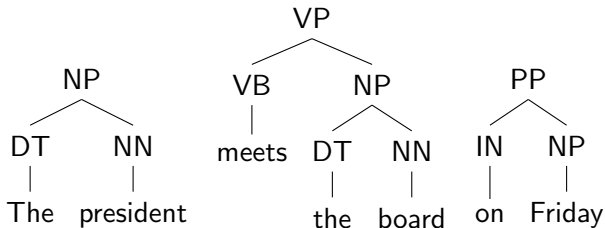
The **president meets the** board on Friday

Parsing



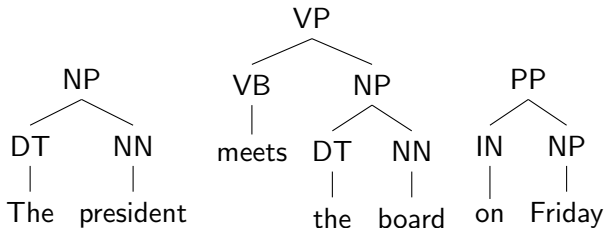
The president **meets the board** on Friday

Parsing



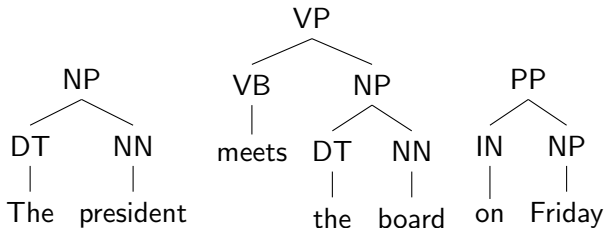
The president meets **the board on** Friday

Parsing



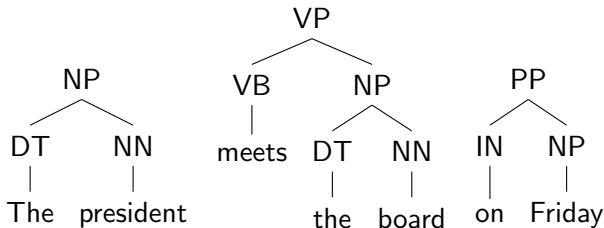
The president meets the **board on Friday**

Parsing



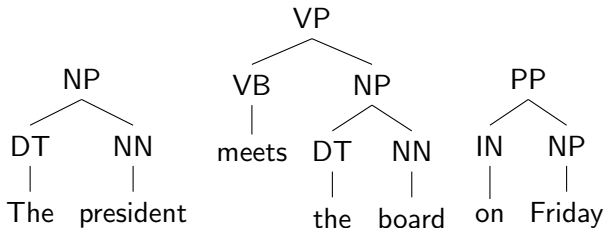
The president meets the board on Friday

Parsing



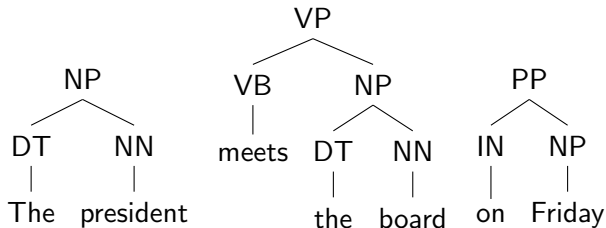
The president meets the board on Friday

Parsing



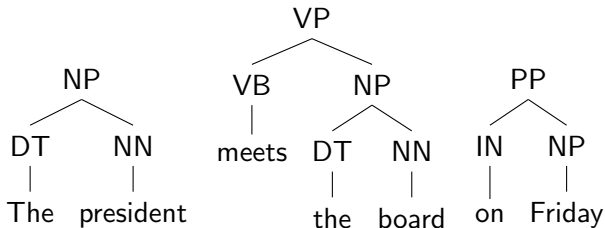
The president **meets the board on** Friday

Parsing



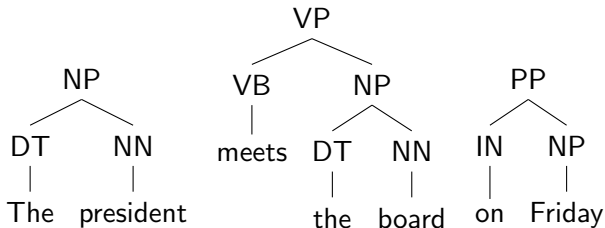
The president meets **the board on Friday**

Parsing

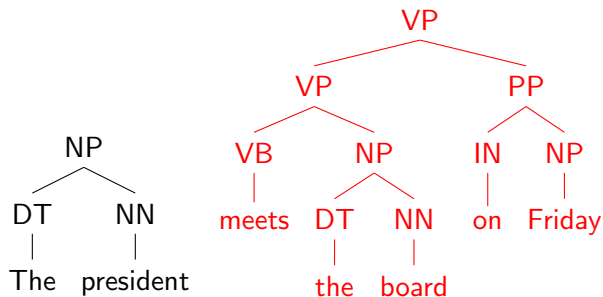


The president meets the board on Friday

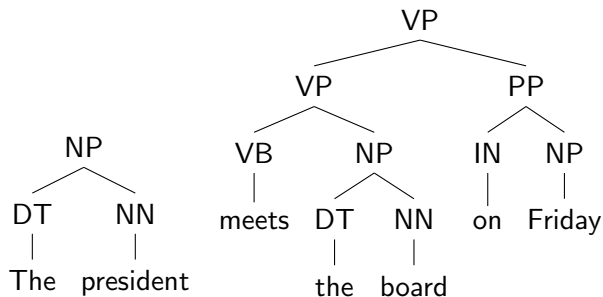
Parsing



The president meets the board on Friday

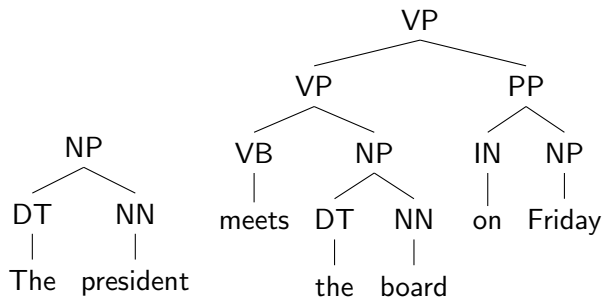


The president **meets the board on Friday**



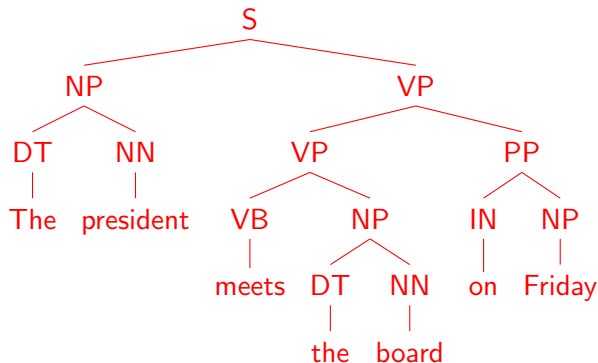
The president meets the board on Friday

Parsing



The president meets the board on Friday

Parsing



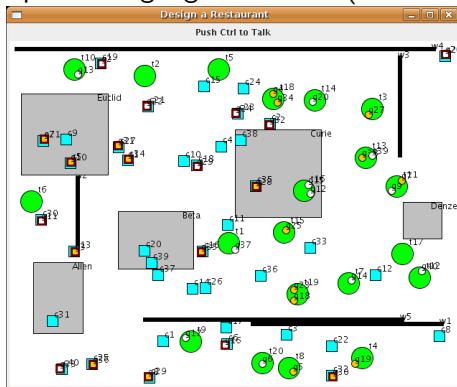
The president meets the board on Friday

Incremental Parsing

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

Incremental Parsing

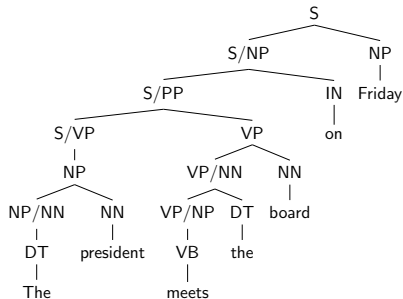
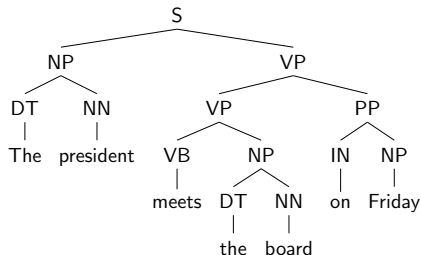
- Spoken language interfaces (Schwartz et al, 2009)



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

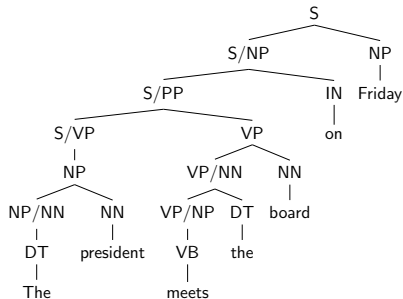
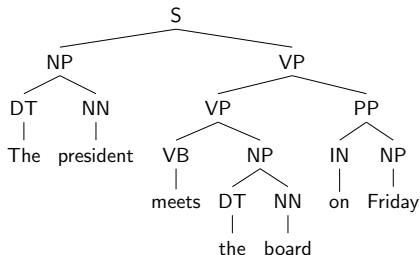
Right-corner Incremental Parsing

Transform right-expanding sequences of constituents



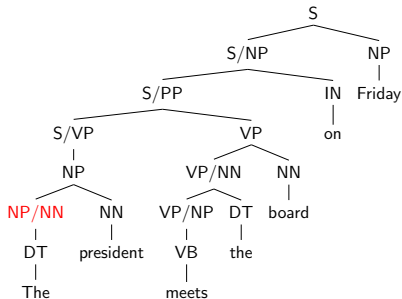
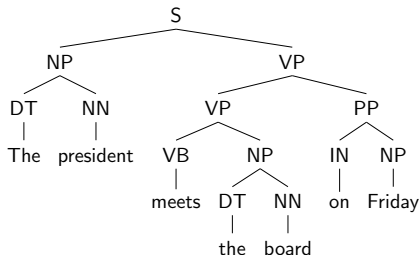
Right-corner Incremental Parsing

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



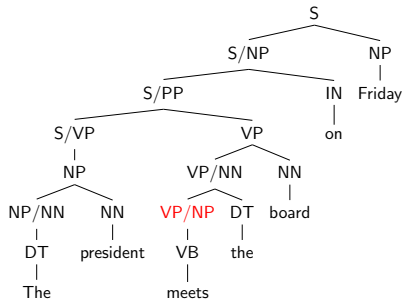
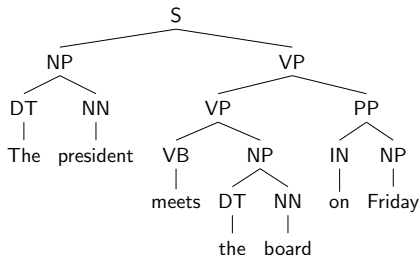
Right-corner Incremental Parsing

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



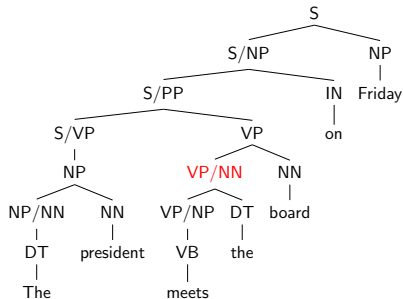
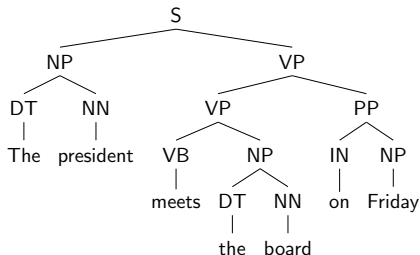
Right-corner Incremental Parsing

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



Right-corner Incremental Parsing

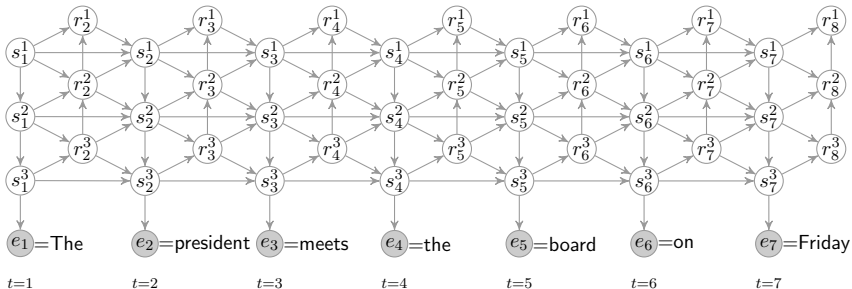
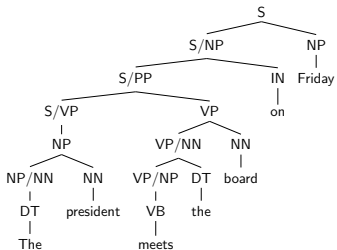
Transform right-expanding sequences of constituents into left-expanding sequences of incomplete constituents



Right-corner Incremental Parsing using HHMM

Hierarchical Hidden Markov Model

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



Motivation
○○○○○○○

Incremental Parsing
○○○○●○

Integration
○○○○○○○○○○

Results

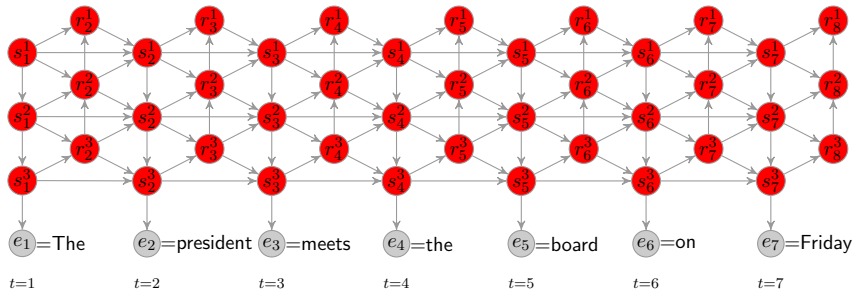
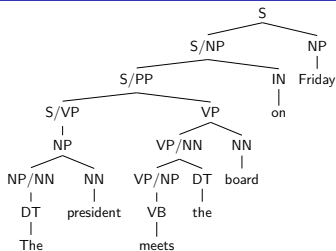
An Incremental Syntactic Language Model for Statistical Phrase-based Machine Translation

Lane Schwartz

Right-corner Incremental Parsing using HHMM

Hierarchical Hidden Markov Model

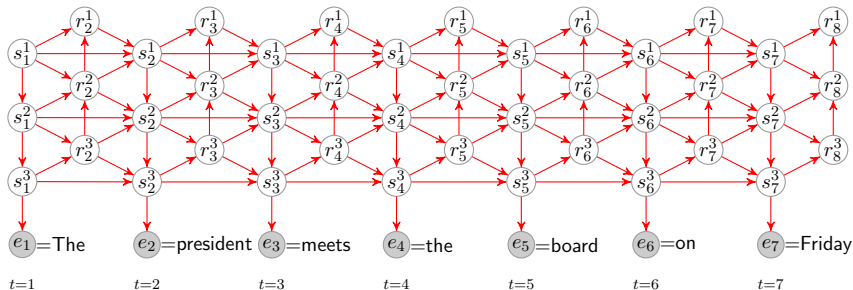
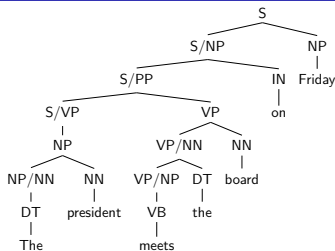
- Circles denote hidden random variables
- Edges denote conditional dependencies
- Shaded circles denote observed values



Right-corner Incremental Parsing using HHMM

Hierarchical Hidden Markov Model

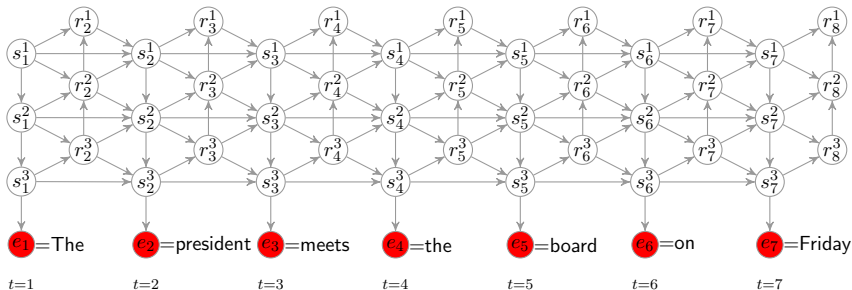
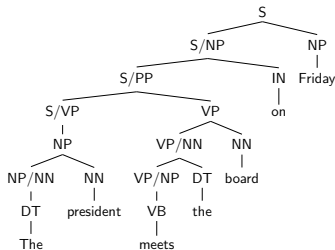
- Circles denote hidden random variables
- Edges denote conditional dependencies
- Shaded circles denote observed values



Right-corner Incremental Parsing using HHMM

Hierarchical Hidden Markov Model

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

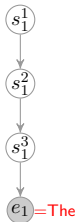
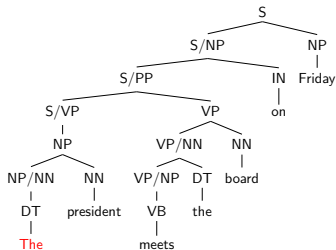


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

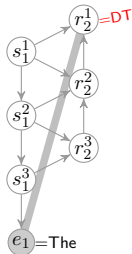
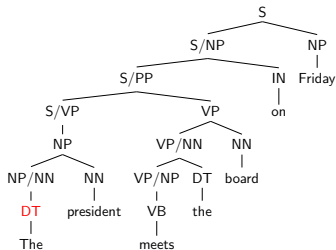


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

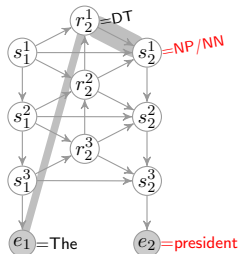
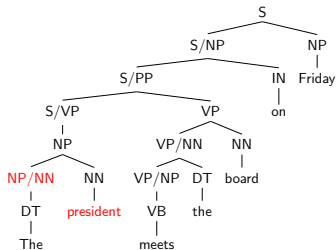


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

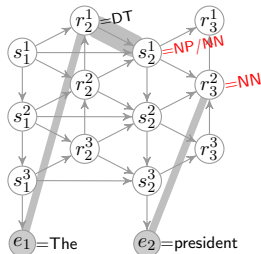
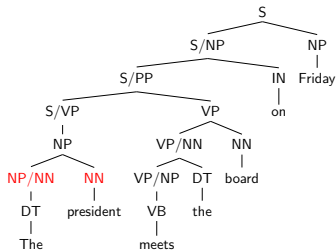


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

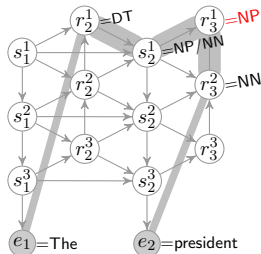
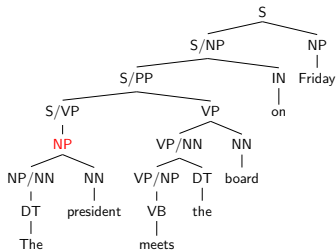


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

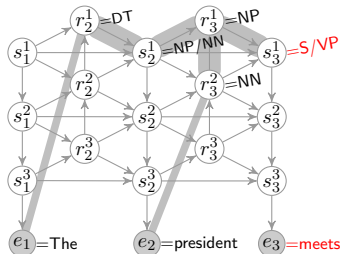
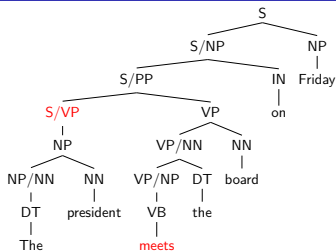


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

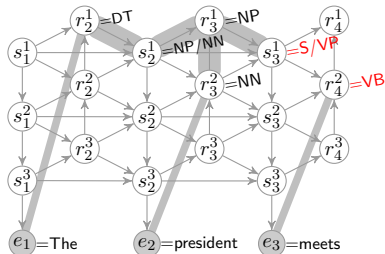
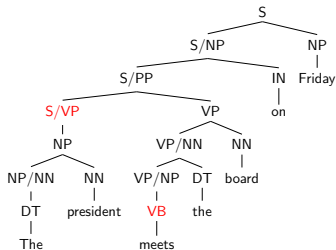


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

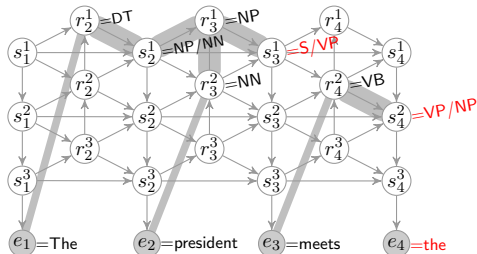
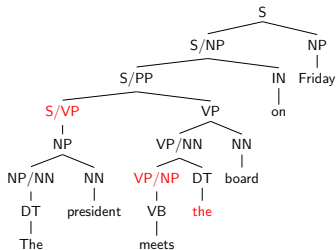


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

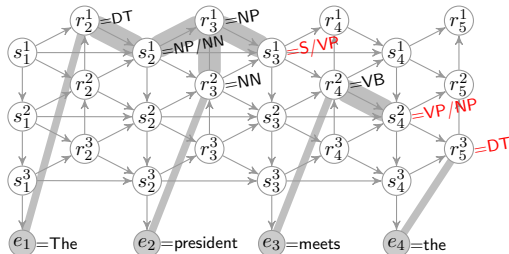
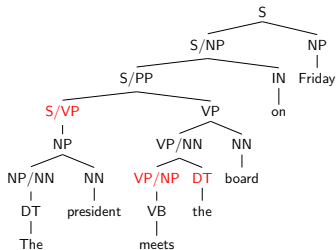


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

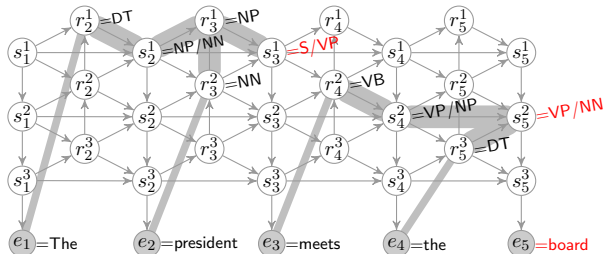
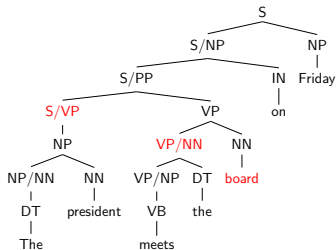


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

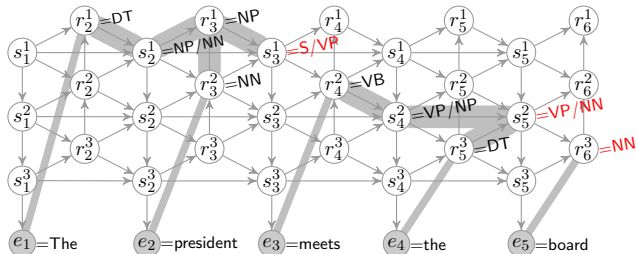
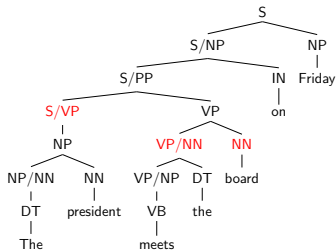


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

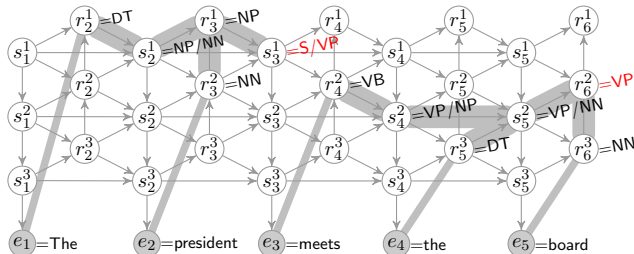
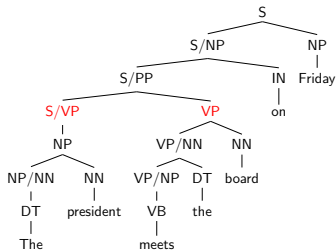


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

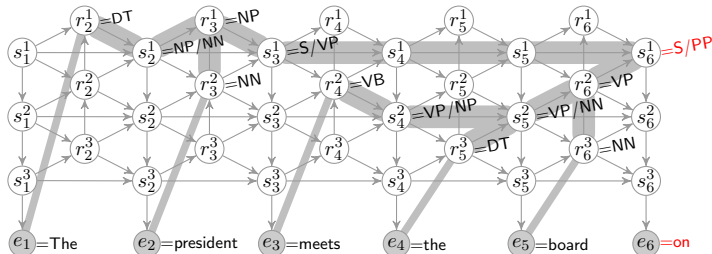
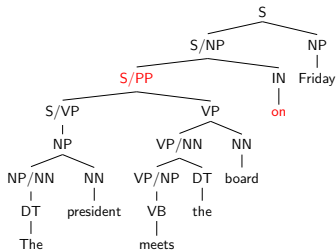


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

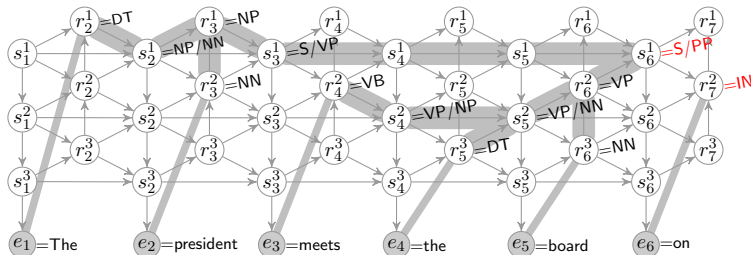
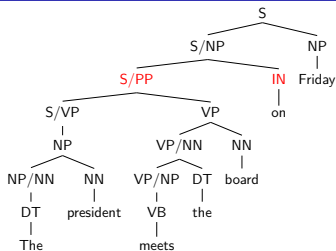


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

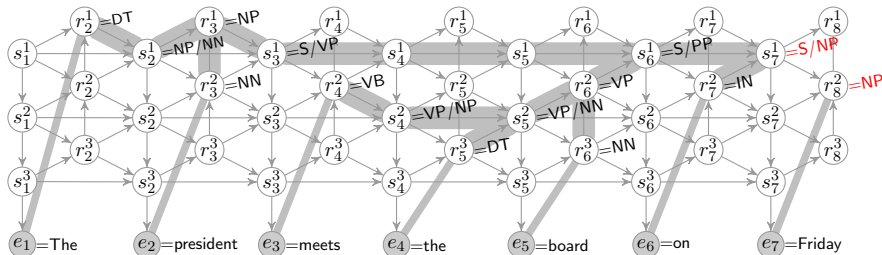
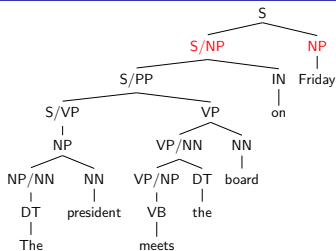


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental”
CCG Parsing

Analogous to Probabilistic
Push-Down Automata

Isomorphic Tree \rightarrow Path

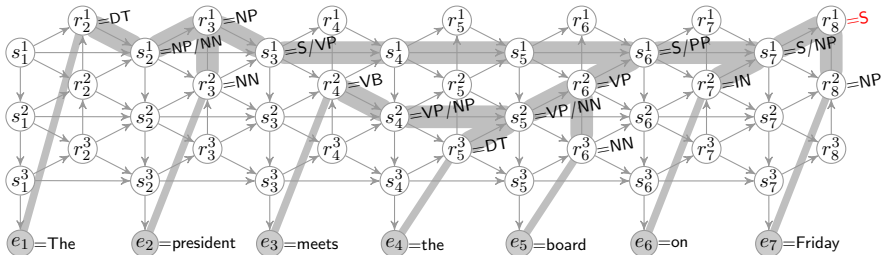
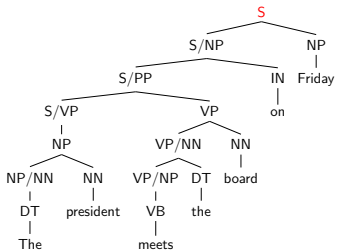


Right-corner Incremental Parsing using HHMM

Analogous to “Maximally Incremental” CCG Parsing

Analogous to Probabilistic Push-Down Automata

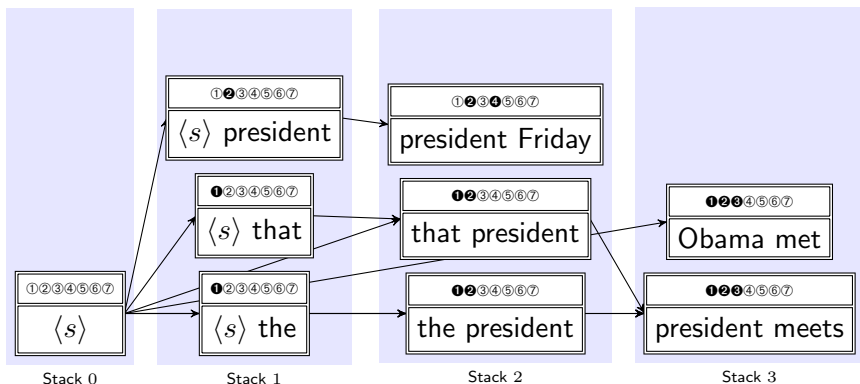
Isomorphic Tree \rightarrow Path



Phrase-Based Translation is also Incremental

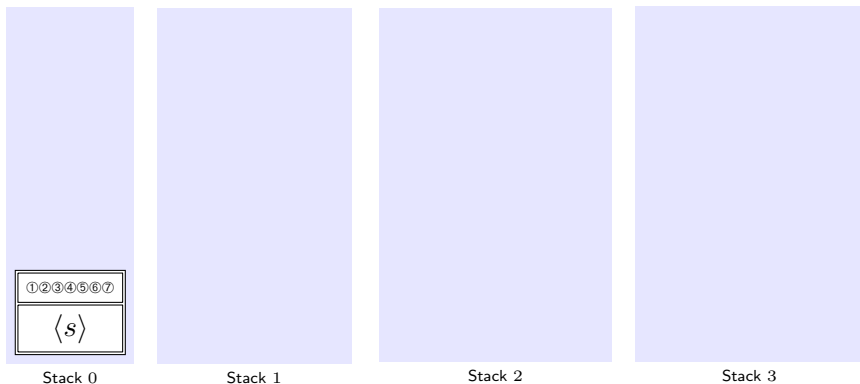
Der Präsident trifft am Freitag den Vorstand

The president meets the board on Friday



Phrase-Based Translation is also Incremental

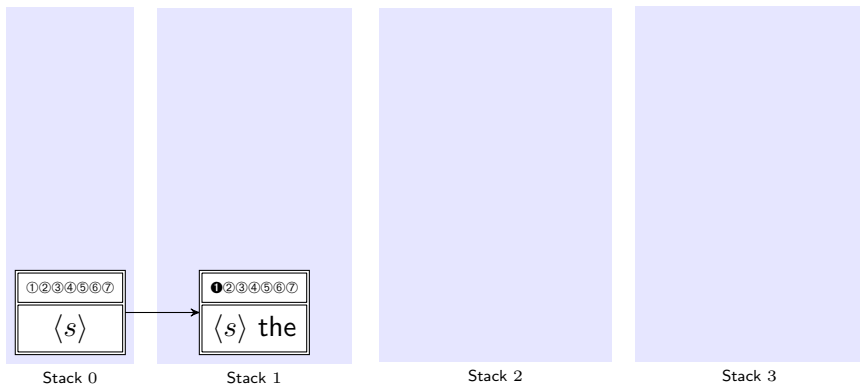
Der Präsident trifft am Freitag den Vorstand



Phrase-Based Translation is also Incremental

Der *Präsident* *trifft* *am* *Freitag* *den* *Vorstand*

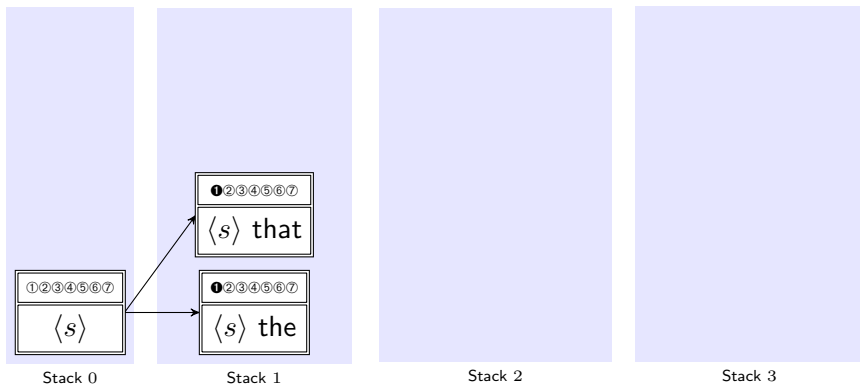
The



Phrase-Based Translation is also Incremental

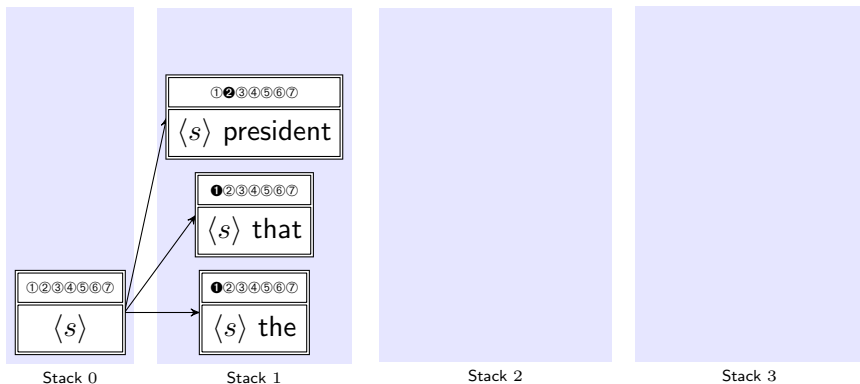
Der Präsident trifft am Freitag den Vorstand

That



Phrase-Based Translation is also Incremental

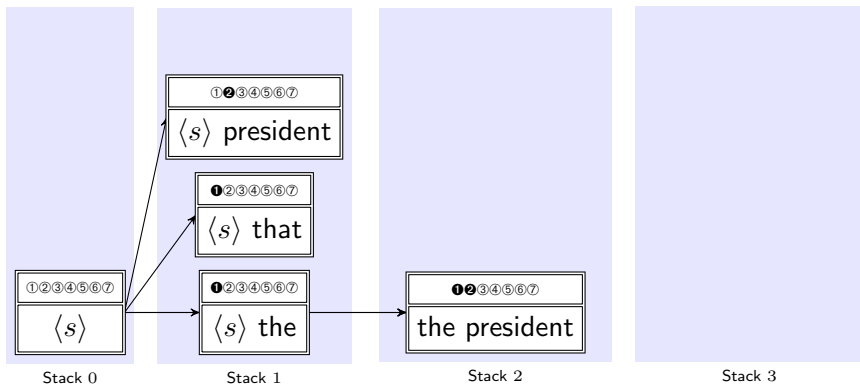
Der *Präsident* trifft am Freitag den Vorstand
President



Phrase-Based Translation is also Incremental

Der Präsident trifft am Freitag den Vorstand

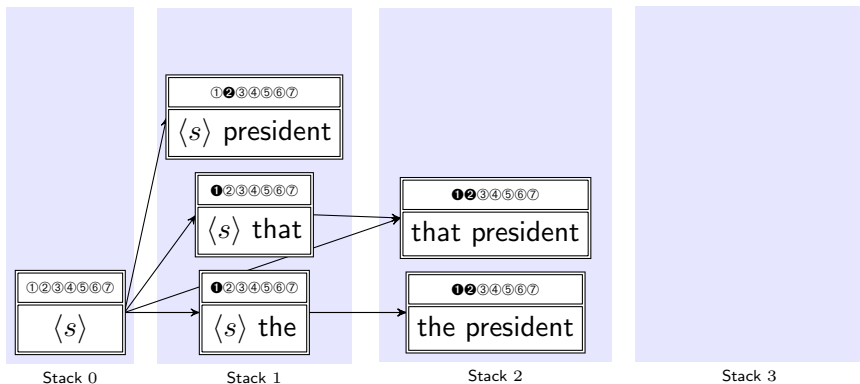
The president



Phrase-Based Translation is also Incremental

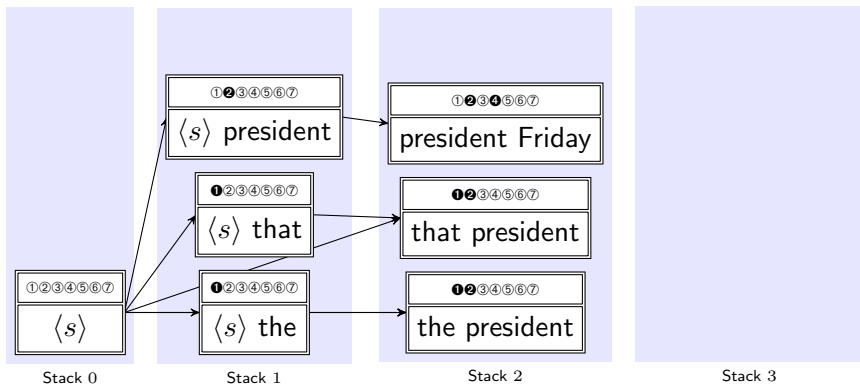
Der Präsident trifft am Freitag den Vorstand

That president



Phrase-Based Translation is also Incremental

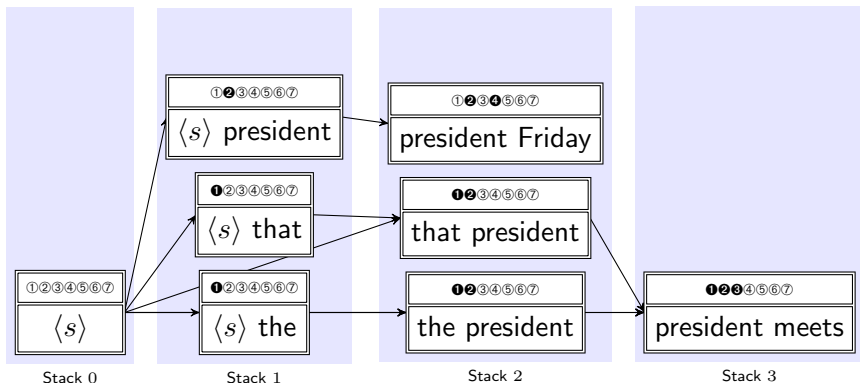
Der *Präsident* trifft am *Freitag* den Vorstand
President Friday



Phrase-Based Translation is also Incremental

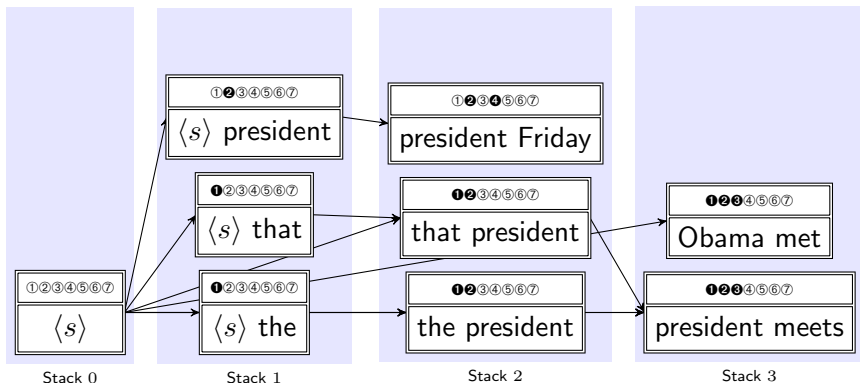
Der *Präsident trifft* am Freitag den Vorstand

The president meets



Phrase-Based Translation is also Incremental

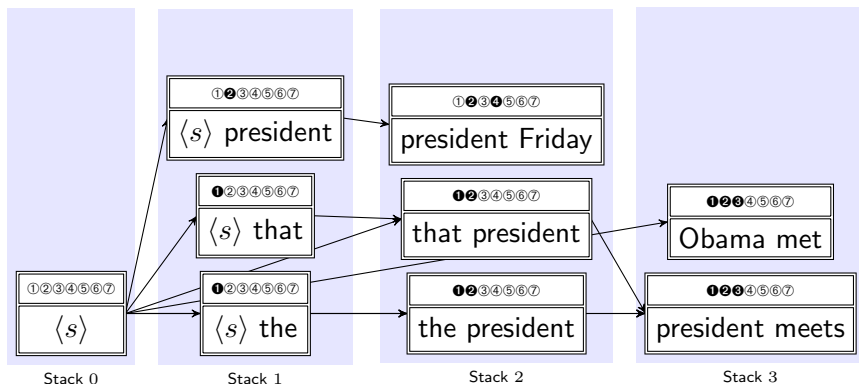
Der *Präsident trifft* am Freitag den Vorstand
Obama met



Phrase-Based Translation is also Incremental

Definition

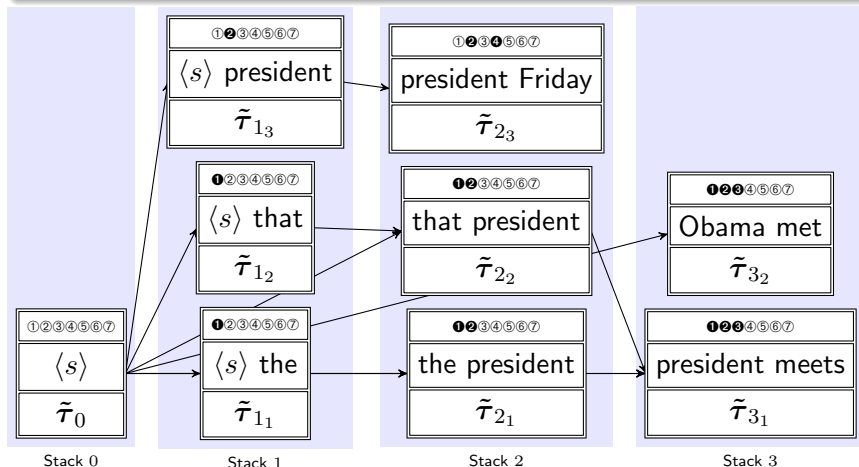
$\tilde{\tau}_{t_h}$ represents parses of the partial translation at node h in stack t



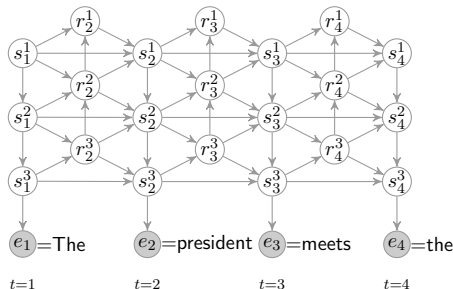
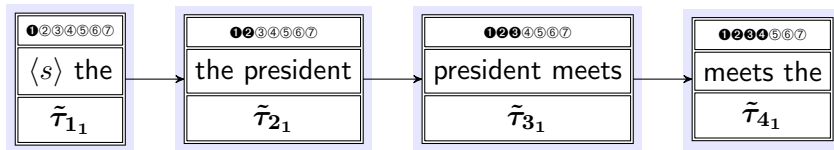
Phrase-Based Translation with Syntactic LM

Definition

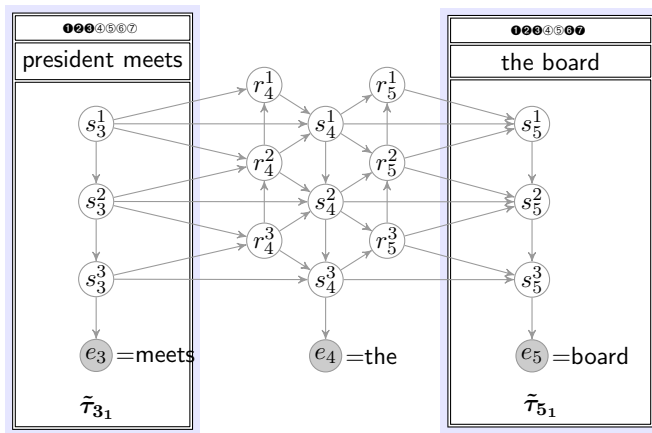
$\tilde{\tau}_{t_h}$ represents parses of the partial translation at node h in stack t



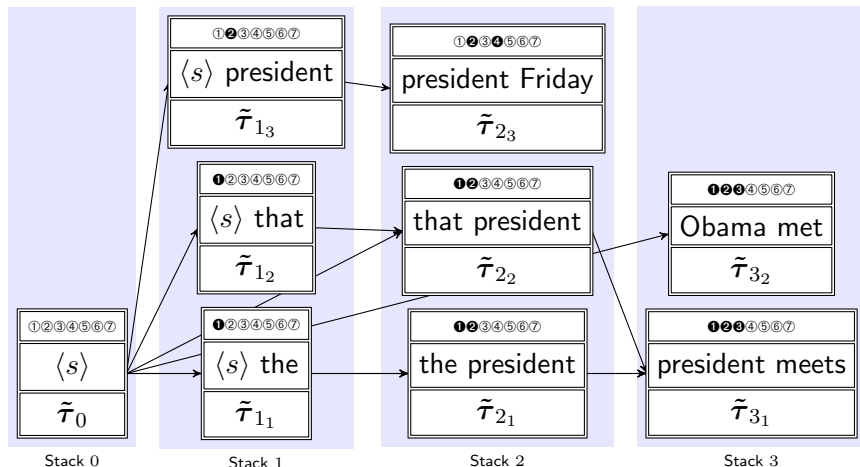
Integrate Parser into Phrase-based Decoder



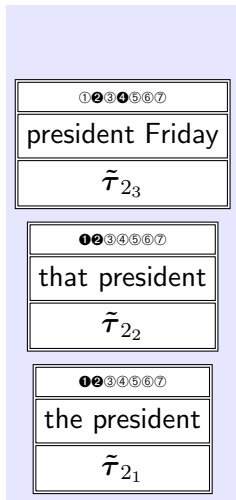
Integrate Parser into Phrase-based Decoder



Syntactic Language Model Guides Pruning



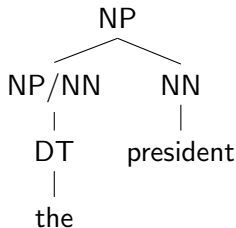
Syntactic Language Model Guides Pruning



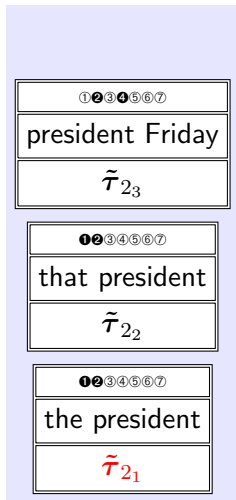
Stack 2

Syntactic Language Model Guides Pruning

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



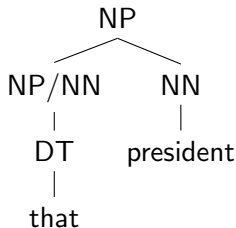
$$P(\tilde{\tau}_{2_1}) = 0.15$$



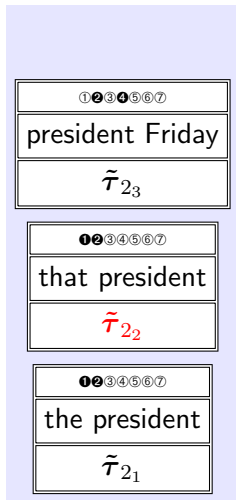
Stack 2

Syntactic Language Model Guides Pruning

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

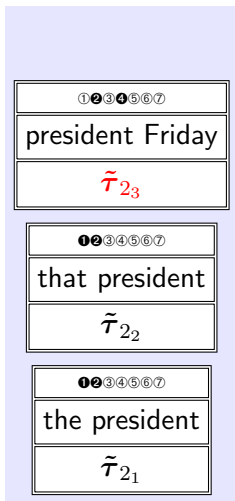


$$P(\tilde{\tau}_{2_2}) = 0.12$$



Stack 2

Syntactic Language Model Guides Pruning



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

NN NN
| |
president Friday

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

Syntactic Language Model Guides Pruning

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

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

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

- BLEU
- Perplexity
- Manual

Experiment

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

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

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

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.

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

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	<u>209</u>	<u>225</u>

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>

Manual Examination

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

Evaluation — Manual

ID	Segment 624, Document "devtest" [$\Delta_{BLEU}=-0.15$]
Source	۔ ' ہر ' 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" [$\Delta_{BLEU}=-0.23$]
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" [$\Delta_{BLEU}=-0.21$]
Source	ہر انسان کو معاشرے میں اپنی 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**

Incremental Parser as Syntactic LM Feature

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

λ = Set of j feature weights

$h = \left\{ \begin{array}{l} \text{Phrase-based translation model} \\ n\text{-gram LM} \\ \text{Distortion model} \\ \vdots \\ \text{Syntactic LM } P(\tilde{\tau}_{t_h}) \end{array} \right.$

Results — Manual Examination

ID	Segment 103, Document "devtest" [$\Delta_{BLEU}=0.68$]
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" [$\Delta_{BLEU}=-0.34$]
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 siddiqui
Reference (reference2)	saiduz zaman siddiqui : a second reference can not be filed against the present chief justice
Reference (reference3)	saiduz zaman siddiqui : 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]