

# COMP 790.139 (Fall 2017)

## Natural Language Processing

### Lecture 7: Summarization; Guest Talk; Machine Translation 1



THE UNIVERSITY  
*of* NORTH CAROLINA  
*at* CHAPEL HILL

Mohit Bansal

(various slides adapted/borrowed from courses by Dan Klein, JurafskyMartin-SLP3, Manning/Socher, others)

# Automatic Document Summarization

# Single-Document Summarization

- ▶ Full document to a salient, non-redundant summary of ~100 words

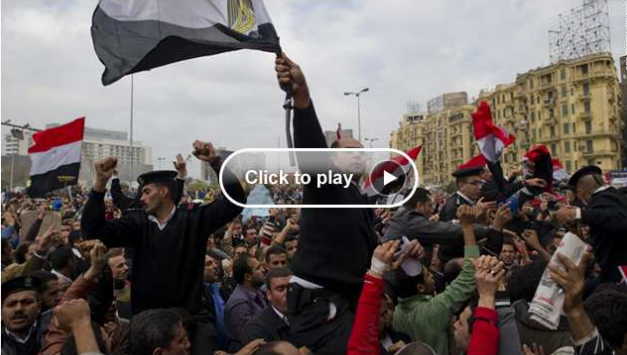
EDITION: U.S. | INTERNATIONAL | MÉXICO  
Set edition preference

CNNWorld

Home Video NewsPulse U.S. World Politics Justice Entertainment Tech Health

## Egypt's military dissolves Parliament, suspends constitution

By the CNN Wire Staff  
February 13, 2011 2:44 p.m. EST



Click to play

Egypt suspends constitution

### STORY HIGHLIGHTS

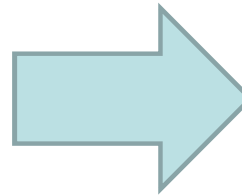
- **NEW:** Banks are shuttered until Wednesday as protests force top banker's resignation
- **NEW:** ElBaradei urges generals to "come out of their headquarters"
- **NEW:** Stock exchange to freeze transactions from officials being investigated
- Egypt's ambassador says the military will run a "technocratic" government until elections

**Cairo, Egypt (CNN)** — Egypt's military dissolved the country's Parliament and suspended its constitution Sunday following the ouster of longtime leader Hosni Mubarak, telling Egyptians it would be in charge for six months or until elections can be held.

The Supreme Council of the Armed Forces said it would appoint a committee to propose changes to the Constitution, which would then be submitted to voters. The council will have the power to issue new laws during the transition period, according to a communique read on state television.

Sameh Shoukry, Egypt's ambassador to the United States, said Sunday that the generals have made restoring security and reviving the economy its top priorities.

"This current composition is basically a technocratic government to run the day-to-day affairs, to take care of the security void that has



### STORY HIGHLIGHTS

- **NEW:** Banks are shuttered until Wednesday as protests force top banker's resignation
- **NEW:** ElBaradei urges generals to "come out of their headquarters"
- **NEW:** Stock exchange to freeze transactions from officials being investigated
- Egypt's ambassador says the military will run a "technocratic" government until elections

# Multi-Document Summarization

- ▶ Several news sources with articles on the same topic (can use overlapping info across articles as a good feature for summarization)

... 27,000+ more

**Egypt's military dissolves parliament**  
By the CNN Wire Staff  
February 13, 2011 2:44 p.m. EST

**Egyptian Military Dissolves Parliament**  
By ANTHONY SHADID  
Published: February 13, 2011

**CAIRO —** The Egyptian military consolidated its control over what it has called a democratic transition from nearly three decades of President Hosni Mubarak's authoritarian rule, dissolving the feeble Parliament, suspending the Constitution and calling for elections in six months in sweeping steps that echoed protesters' demands.

**STORY HIGHLIGHTS**

- **NEW:** Banks are shuttered until Wednesday as protests force top banker's resignation
- **NEW:** ElBaradei urges generals to "come out of their headquarters"
- **NEW:** Stock exchange to freeze transactions from officials being investigated
- Egypt's ambassador says the military will run a "technocratic" government until elections

**Cairo, Egypt** (C) Parliament and ouster of longtime be in charge for

The Supreme Council committee to be submitted to laws during the on state televis

Samah Shoukry, Egypt's ambassador to the United States, said Sunday that the generals have made restoring security and reviving the economy its top priorities.

"This current composition is basically a technocratic government to run the day-to-day affairs, to take care of the security void that has

**Only at BlackBerry**  
Get the ScoreMobile Super App

**Massive Population Lifts Nation's Growth**

**Mideast Unrest Spreads**  
Protests Target Iran, Bahrain, Libya; Egypt Dissolves Parliament

**Article**  
By MARGARET COCKER, MATT BRADLEY and TAMER EL

**IDEOLGY shadows fight to rule**  
By Oussama Dine, USA TODAY

**CAIRO —** As Abdel Fattah, spokesman for the long-banned Islamist group, announced his office in South Cairo, answering questions in dark suits and discussing plans for Egypt under a new

**Twitter**  
SIGN IN TO E-MAIL  
PRINT  
SINGLE PAGE  
REPRINTS  
SHARE  
NOW PLAYING EVERYWHERE

# Extractive Summarization

---

- ▶ Directly selecting existing sentences from input document instead of rewriting them

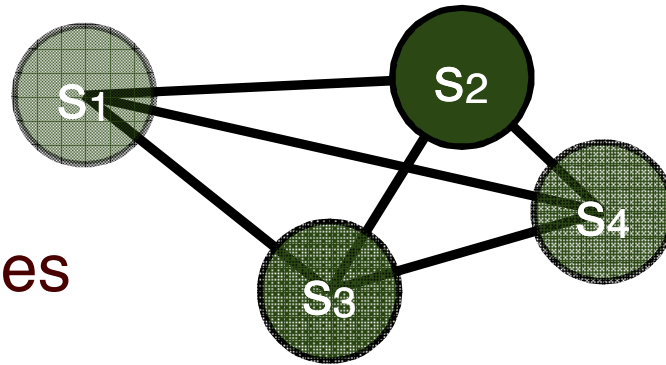
- S<sub>1</sub>** The health care bill is a major test for the Obama administration.
- S<sub>2</sub>** Universal health care is a divisive issue.
- S<sub>3</sub>** President Obama remained calm.
- S<sub>4</sub>** Obama addressed the House on Tuesday.

# Graph-based Extractive Summ

---

Stationary distribution  
represents node centrality

Nodes are sentences



Edges are similarities

# Maximize Concept Coverage

- S<sub>1</sub>** The health care bill is a major test for the Obama administration.
- S<sub>2</sub>** Universal health care is a divisive issue.
- S<sub>3</sub>** President Obama remained calm.
- S<sub>4</sub>** Obama addressed the House on Tuesday.

concept	value
obama	3
health	2
house	1

Length limit:  
18 words

summary	length	value
{S <sub>1</sub> , S <sub>3</sub> }	17	5
{S <sub>2</sub> , S <sub>3</sub> , S <sub>4</sub> }	17	6

← greedy

← optimal

# Maximize Concept Coverage

- ▶ A set coverage optimization problem

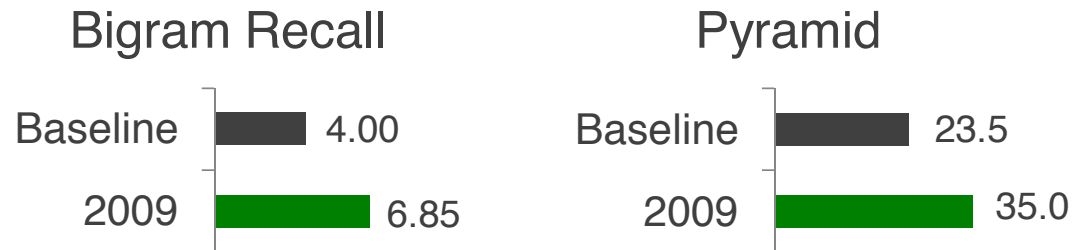
$$\max_{s \in S(D)} \sum_{c \in C(s)} v_c$$

Set of extractive summaries of document set D

Value of concept c

Set of concepts present in summary s

## Results








# Maximize Concept Coverage

---

- ▶ Can be solved using an integer linear program with constraints:

Maximize:  $\sum_i w_i c_i$   total concept value

Subject to:  $\sum_j l_j s_j \leq L$   summary length limit

$s_j Occ_{ij} \leq c_i, \quad \forall i, j$   
 $\sum_j s_j Occ_{ij} \geq c_i \quad \forall i$   maintain consistency between  
selected sentences and concepts

$c_i \in \{0, 1\} \quad \forall i$

$s_j \in \{0, 1\} \quad \forall j$

# Beyond Extraction: Compression

---

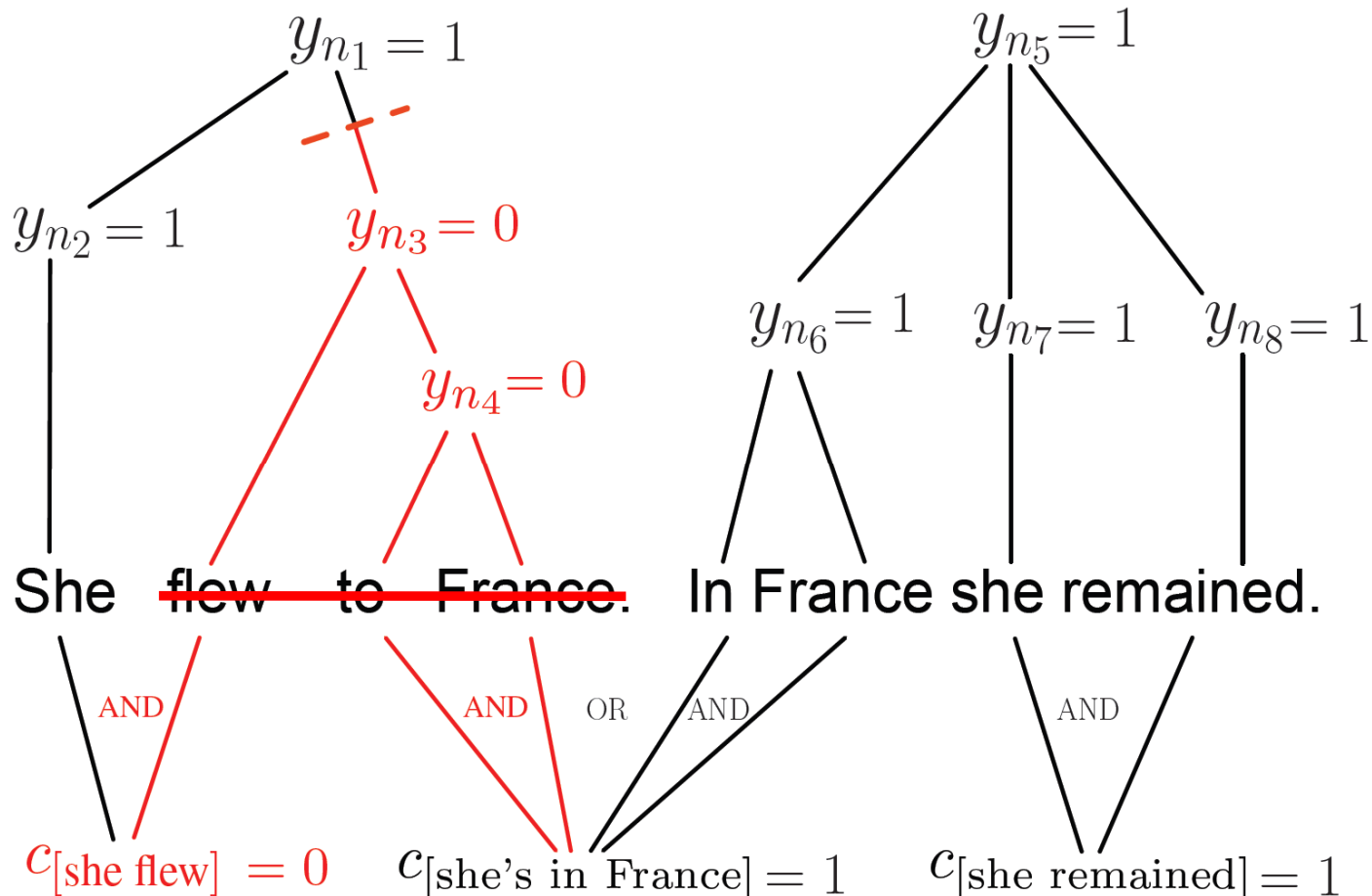
- ▶ If you had to write a concise summary, making effective use of the 100-word limit, you would remove some information from the lengthy sentences in the original article

What would a human do?

~~It is therefore unsurprising that~~ Lindsay pleaded not guilty ~~yesterday afternoon~~ to the charges filed against her, ~~according to her publicist.~~

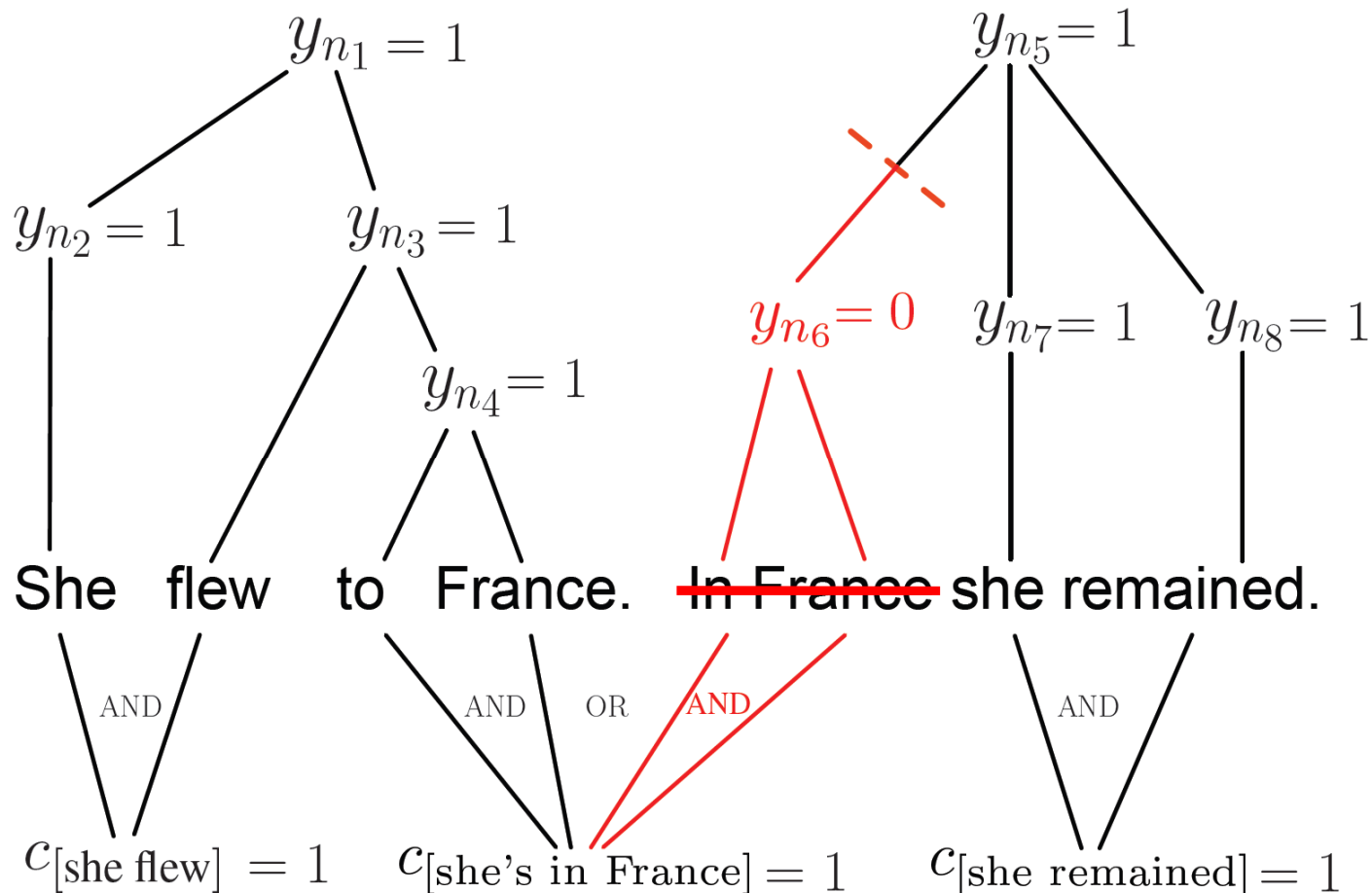
# Beyond Extraction: Compression

- ▶ Model should learn the subtree deletions/cuts that allow compression



# Beyond Extraction: Compression

- ▶ Model should learn the subtree deletions/cuts that allow compression



# Beyond Extraction: Compression

- ▶ The new optimization problem looks to maximize the concept values as well as safe deletion values in the candidate summary:

$$\max_{s \in S(D)} \left[ \sum_{c \in C(s)} v_c + \sum_{d \in D(s)} v_d \right]$$

Value of deletion d

Set branch cut deletions made in creating summary s

- ▶ To decide the value/cost of a deletion, we decide relevant deletion features and the model learns their weights:

$$v_d = w^\top f(d)$$

# Beyond Extraction: Compression

---

- ▶ Some example features for concept bigrams and cuts/deletions:

## Bigram Features $f(b)$

COUNT:	Bucketed document counts
STOP:	Stop word indicators
POSITION:	First document position indicators
CONJ:	All two- and three-way conjunctions of above
BIAS:	Always one

## Cut Features $f(c)$

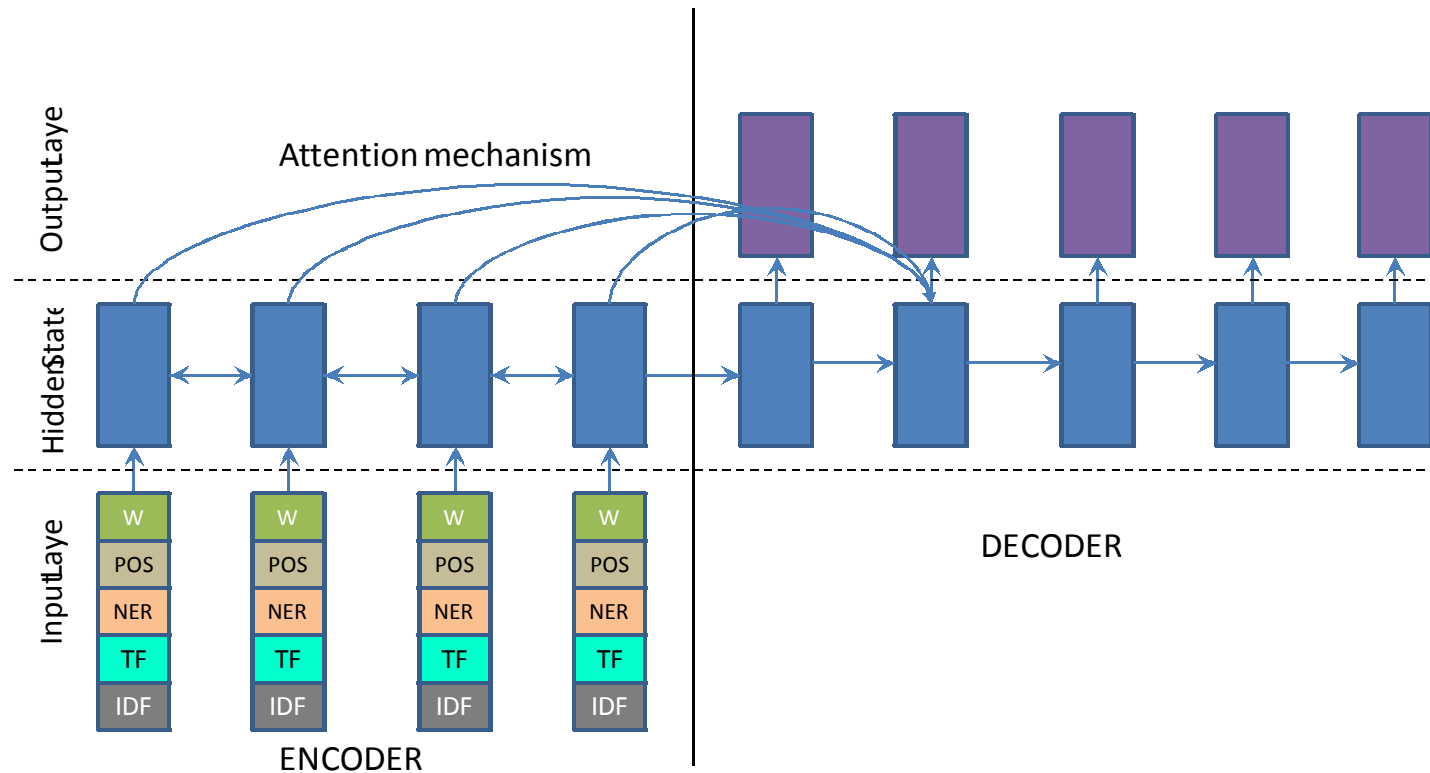
COORD:	Coordinated phrase, four versions: NP, VP, S, SBAR
S-ADJUNCT:	Adjunct to matrix verb, four versions: CC, PP, ADVP, SBAR
REL-C:	Relative clause indicator
ATTR-C:	Attribution clause indicator
ATTR-PP:	PP attribution indicator
TEMP-PP:	Temporal PP indicator
TEMP-NP	Temporal NP indicator
BIAS:	Always one

# Neural Abstractive Summarization

---

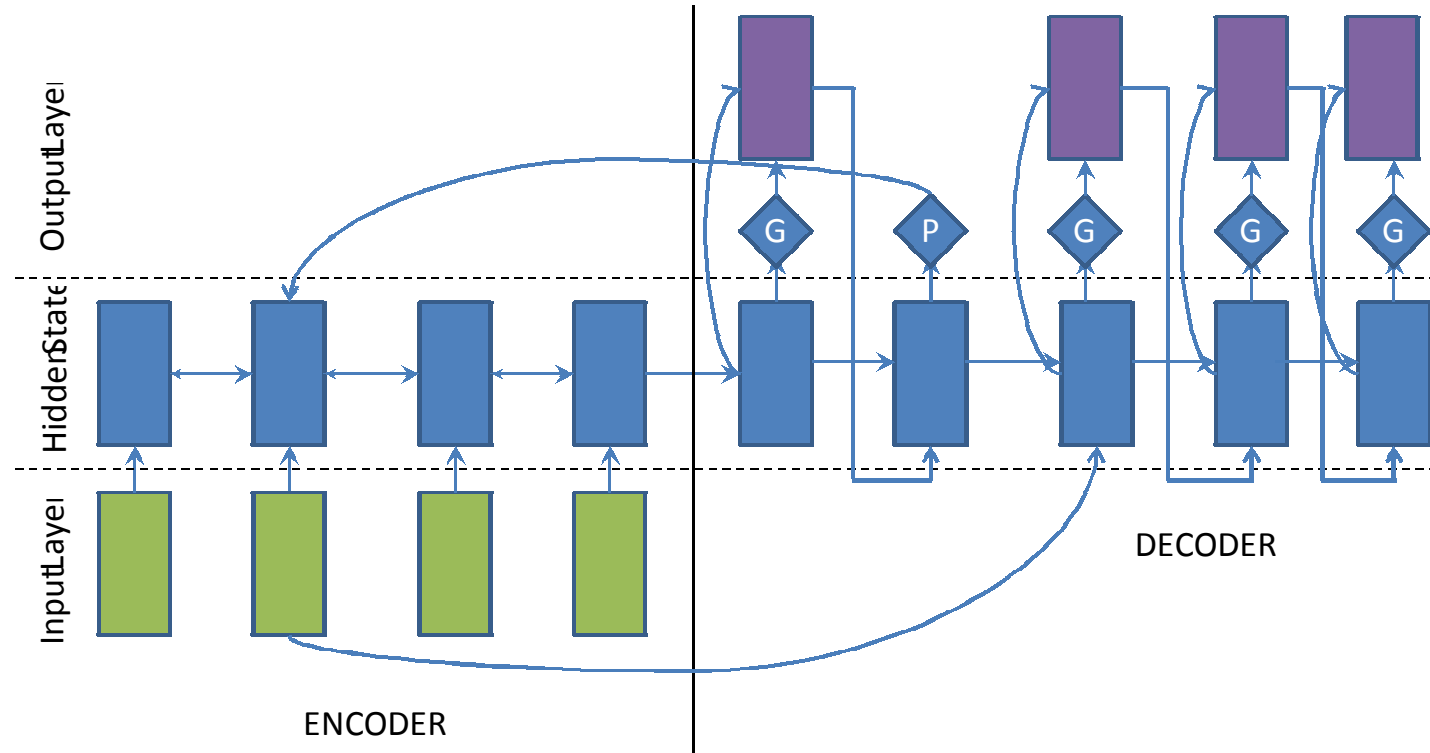
- ▶ Mostly based on sequence-to-sequence RNN models
- ▶ Later added attention, coverage, pointer/copy, hierarchical encoder/attention, metric rewards RL, etc.
- ▶ Examples: Rush et al., 2015; Nallapati et al., 2016; See et al., 2017; Paulus et al., 2017

# Feature-Augmented Encoder-Decoder

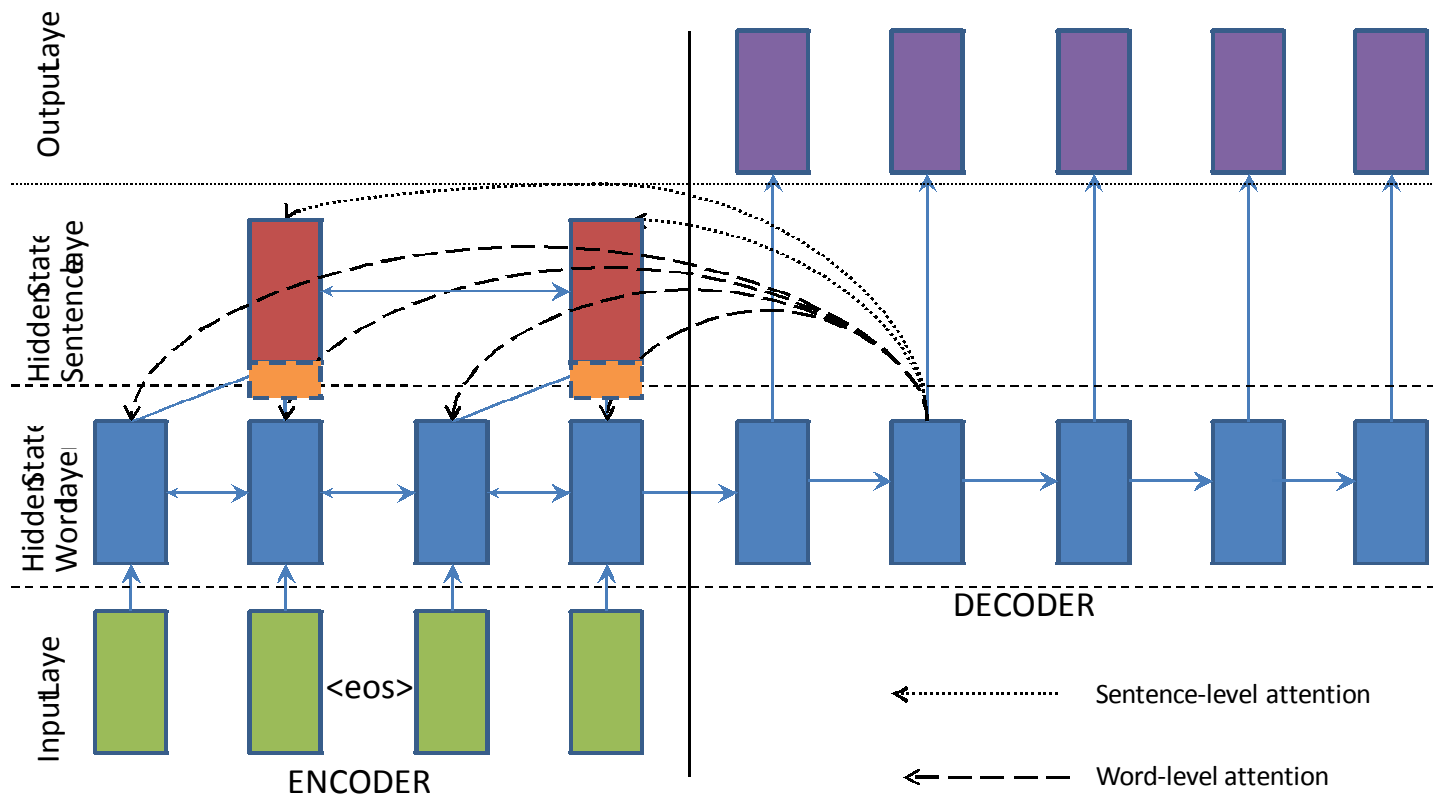




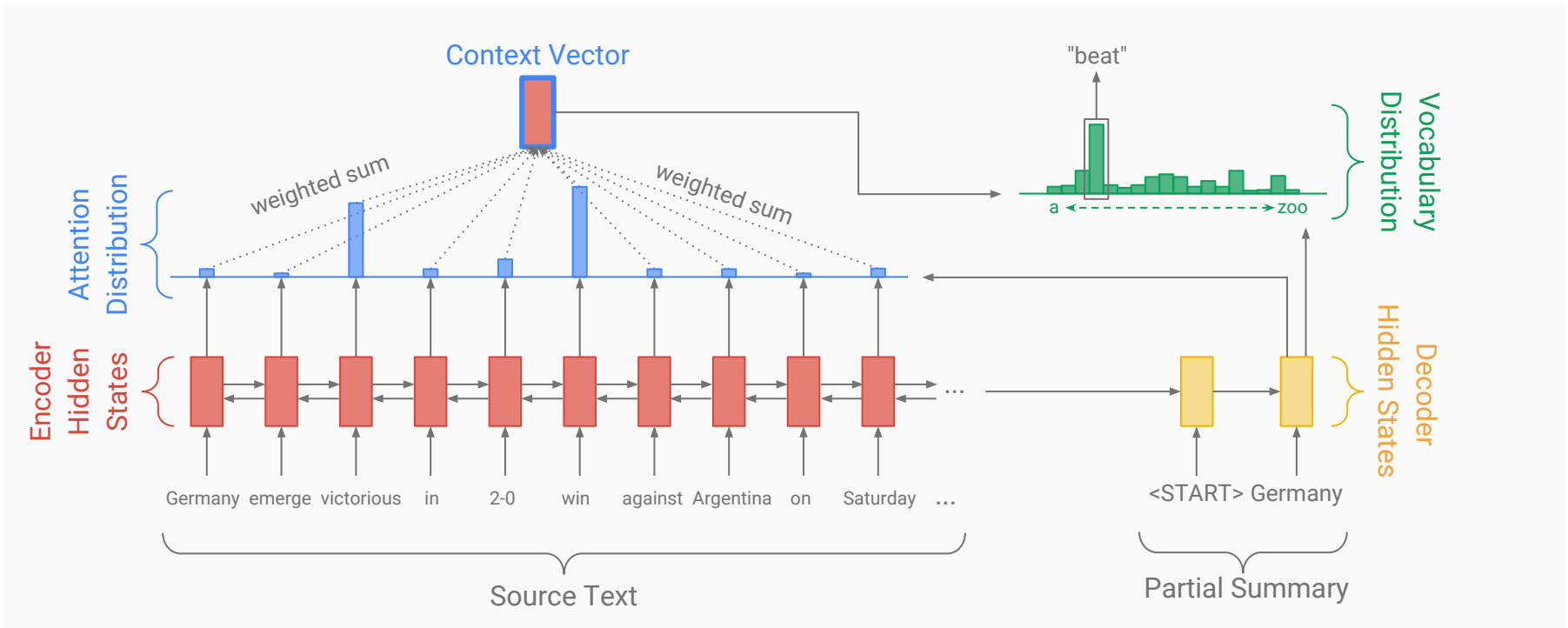
# Generation+Copying



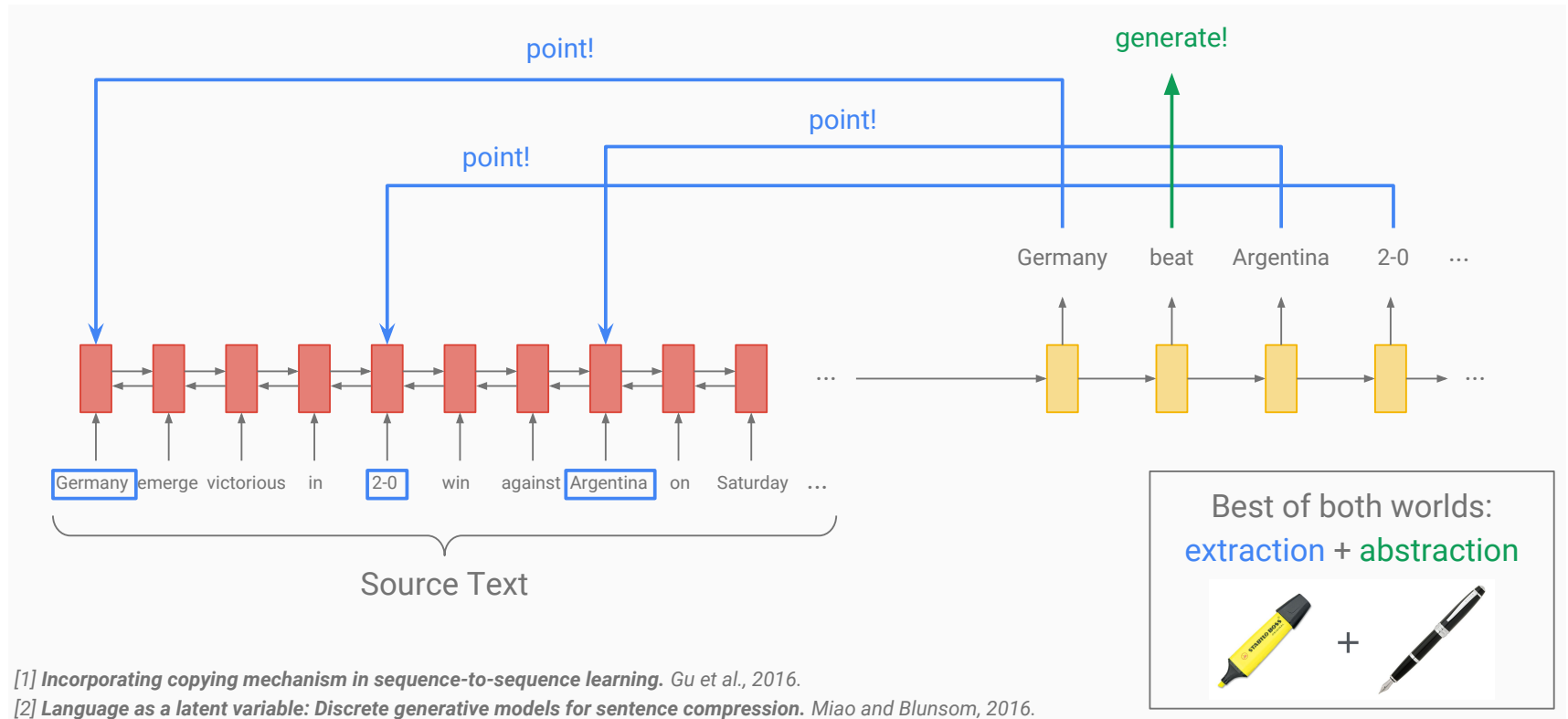
# Hierarchical Attention



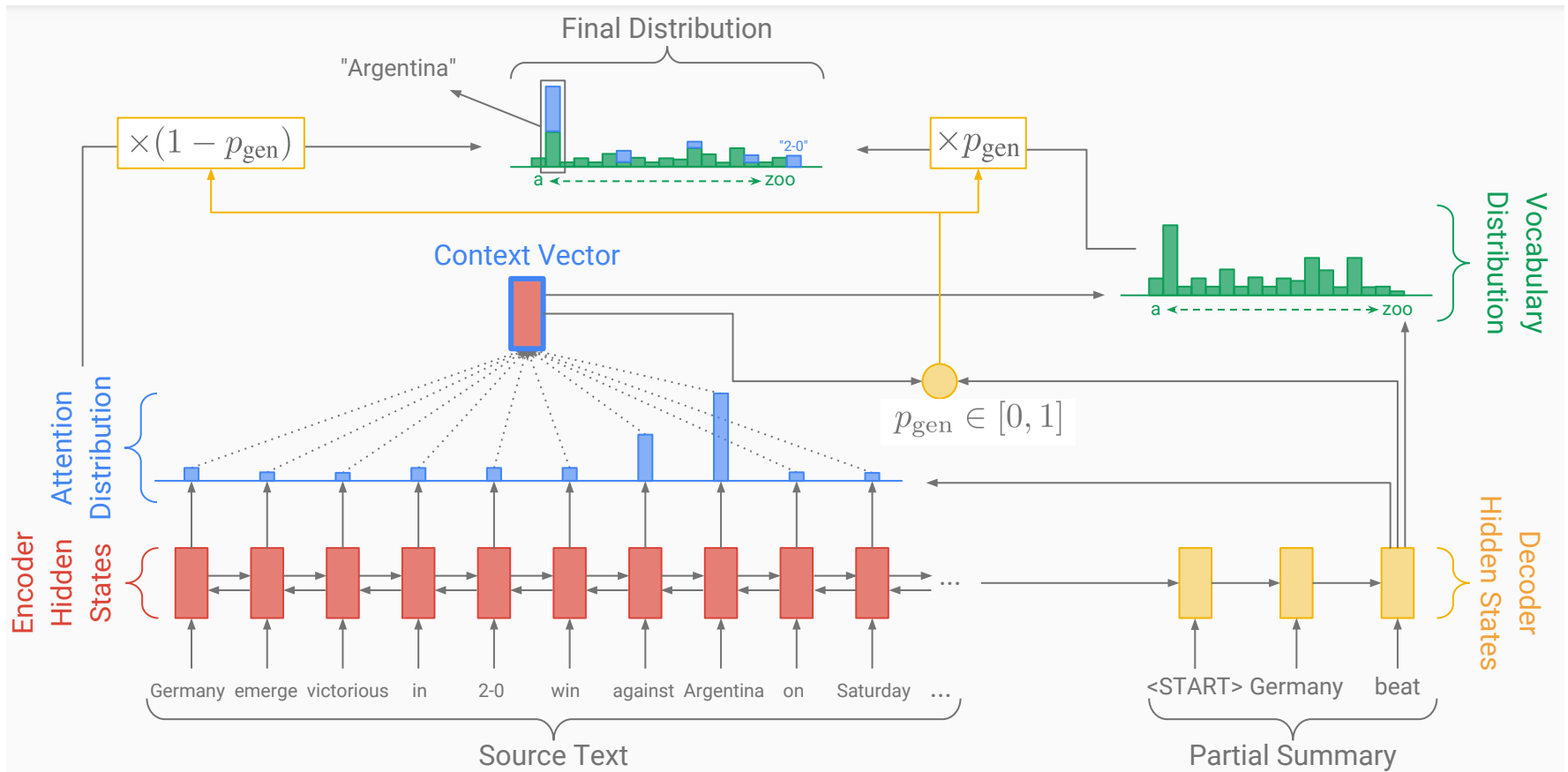
# Pointer-Generator Networks



# Pointer-Generator Networks

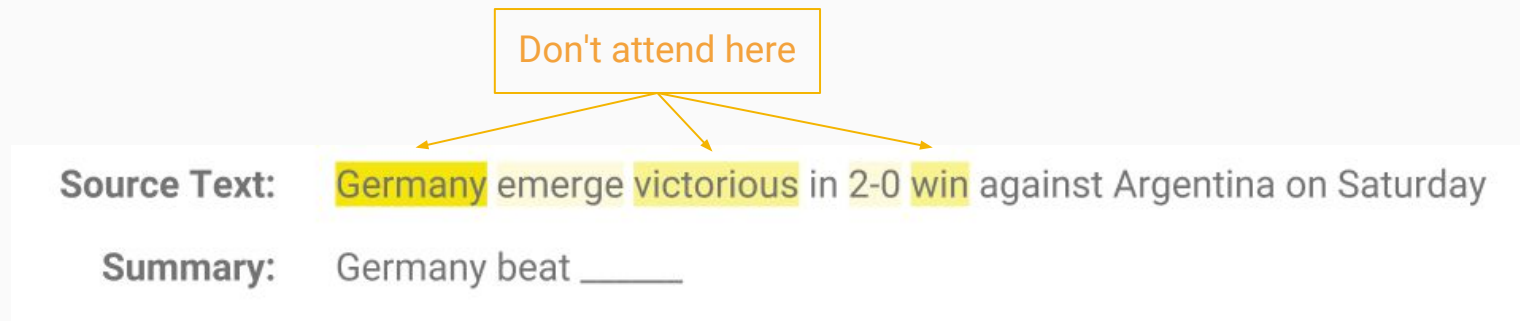


# Pointer-Generator Networks



# Coverage for Redundancy Reduction

**Coverage** = cumulative attention = what has been covered so far



1. Use coverage as **extra input to attention mechanism**.
2. **Penalize** attending to things that have already been covered.

**Result:** repetition rate reduced to level similar to human summaries

[4] *Modeling coverage for neural machine translation*. Tu et al., 2016,

[5] *Coverage embedding models for neural machine translation*. Mi et al., 2016

[6] *Distraction-based neural networks for modeling documents*. Chen et al., 2016.

Guest Talk by Ramakanth Pasunuru:

“Towards Improving Abstractive Summarization via  
Entailment Generation”


(30 mins)





# Machine Translation




# Machine Translation


- ▶ Useful for tons of companies, online traffic, and our international communication!



+Mohit    



Translate 






Hindi English Spanish Detect language ▼



English Spanish Hindi ▼

Translate

This is an example of machine<sup>x</sup> translation  
 

यह मशीन अनुवाद का एक उदाहरण है  
    

Yaha maśīna anuvāda kā ēka udāharaṇa hai

# Statistical Machine Translation

---

- ▶ Source language  $f$  (e.g., French)
- ▶ Target language  $e$  (e.g., English)
- ▶ We want the best target (English) translation given the source (French) input sentence, hence the probabilistic formulation is:

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

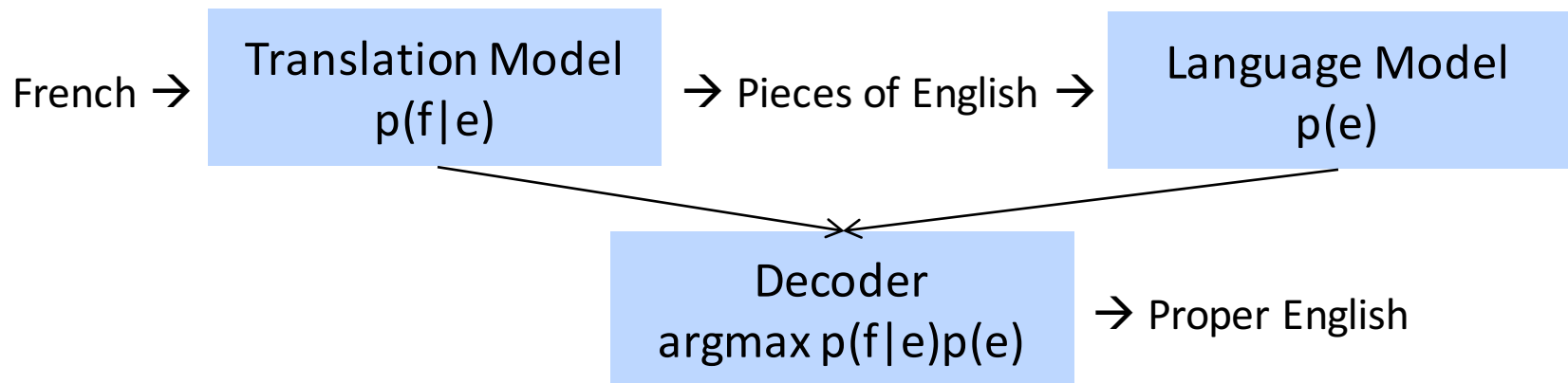
- ▶ Using Bayes rule, we get the following (since  $p(f)$  in the denominator is independent of the  $\operatorname{argmax}$  over  $e$ ):

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

# Statistical Machine Translation

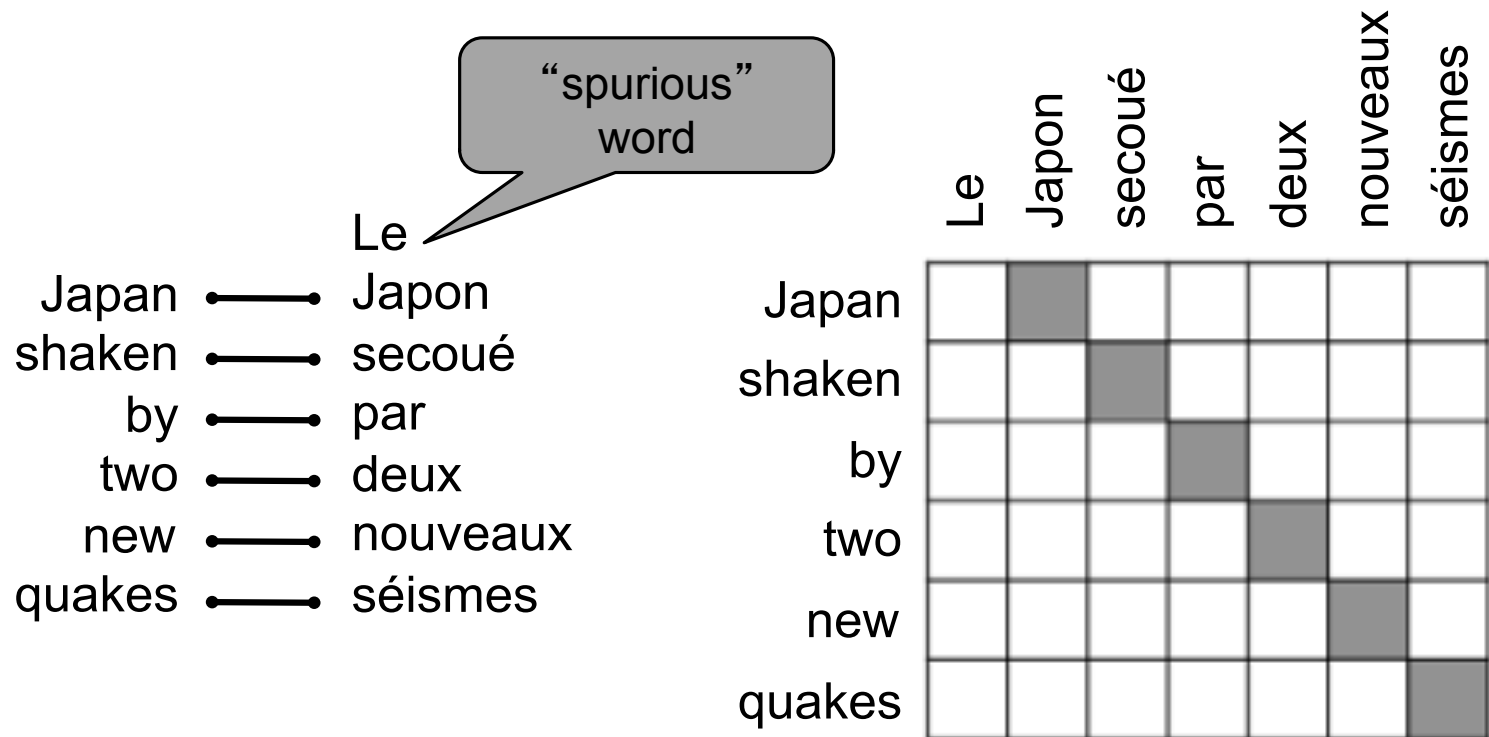
---

- ▶ The first part is known as the 'Translation Model'  $p(f|e)$  and is trained on parallel corpora of  $\{f,e\}$  sentence pairs, e.g., from EuroParl or Canadian parliament proceedings in multiple languages
- ▶ The second part  $p(e)$  is the 'Language Model' and can be trained on tons more monolingual data, which is much easier to find!



# Statistical Machine Translation

- ▶ First step in traditional machine translation is to find alignments or translational matchings between the two sentences, i.e., predict which words/phrases in French align to which words/phrases in English.
- ▶ Challenging problem: e.g., some words may not have any alignments:



# Statistical Machine Translation

- ▶ One word in the source sentence might align to several words in the target sentence:

“zero fertility” word  
not translated

And the program has been implemented

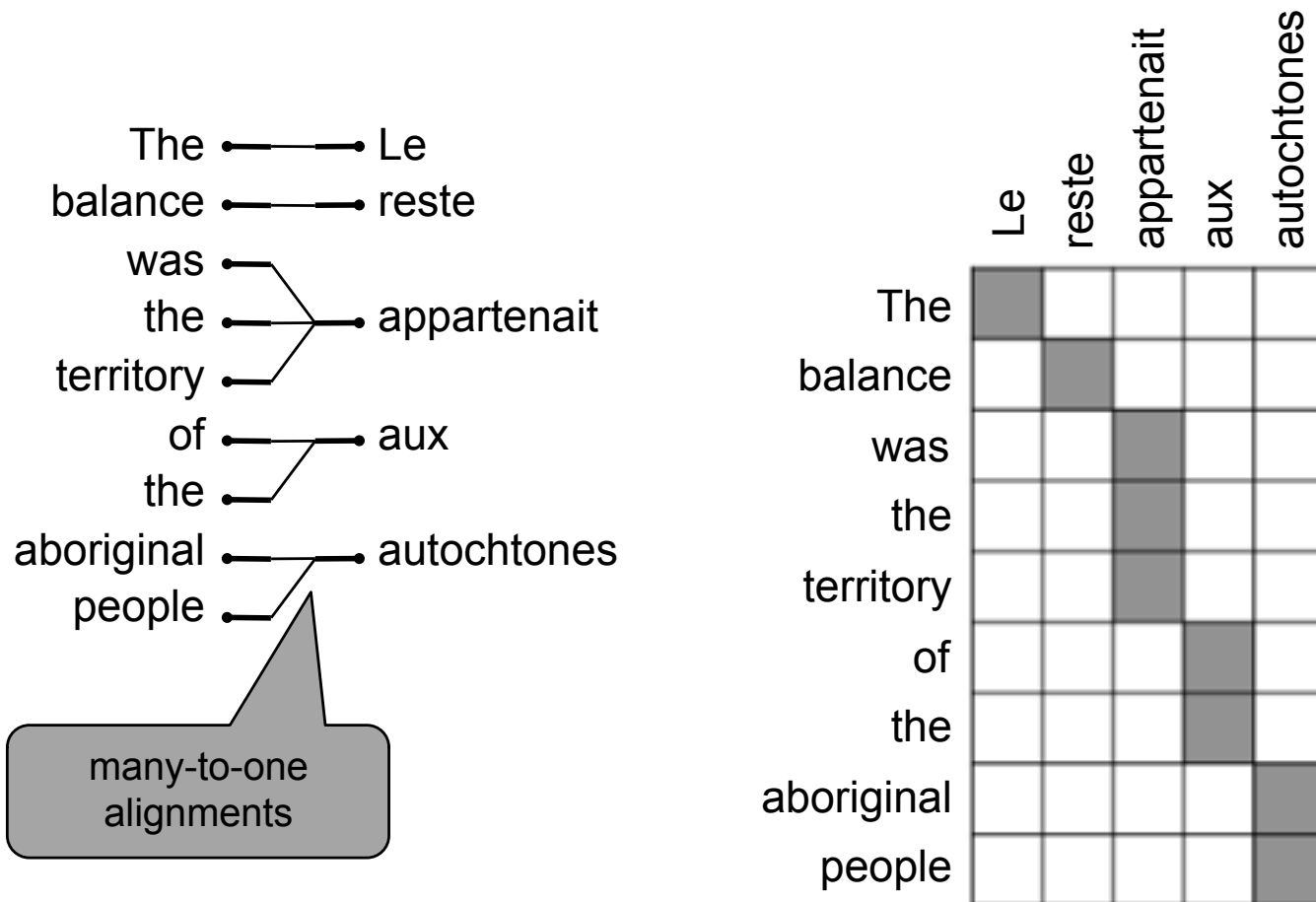
Le programme a été mis en application

one-to-many  
alignment

	Le	programme	a	été	mis	en	application
And							
the							
program							
has							
been							
implemented							

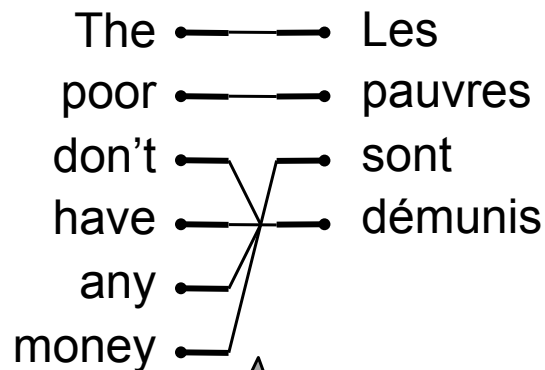
# Statistical Machine Translation

- ▶ Many words in the source sentence might align to a single word in the target sentence:



# Statistical Machine Translation

- ▶ And finally, many words in the source sentence might align to many words in the target sentence:



many-to-many  
alignment

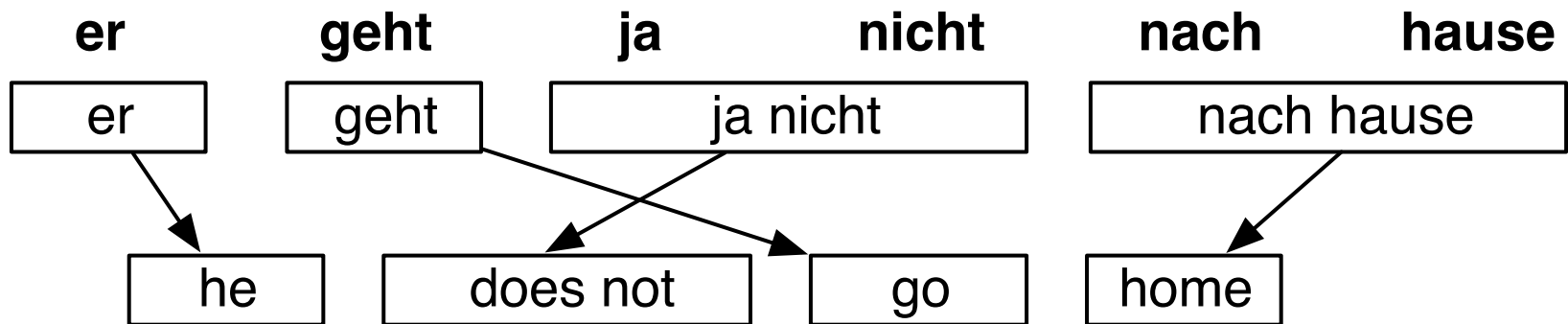
	Les	pauvres	sont	démunis
The				
poor				
don't				
have				
any				
money				

phrase  
alignment

# Statistical Machine Translation

---

- ▶ After learning the word and phrase alignments, the model also needs to figure out the reordering, esp. important in language pairs with very different orders!





# Statistical Machine Translation

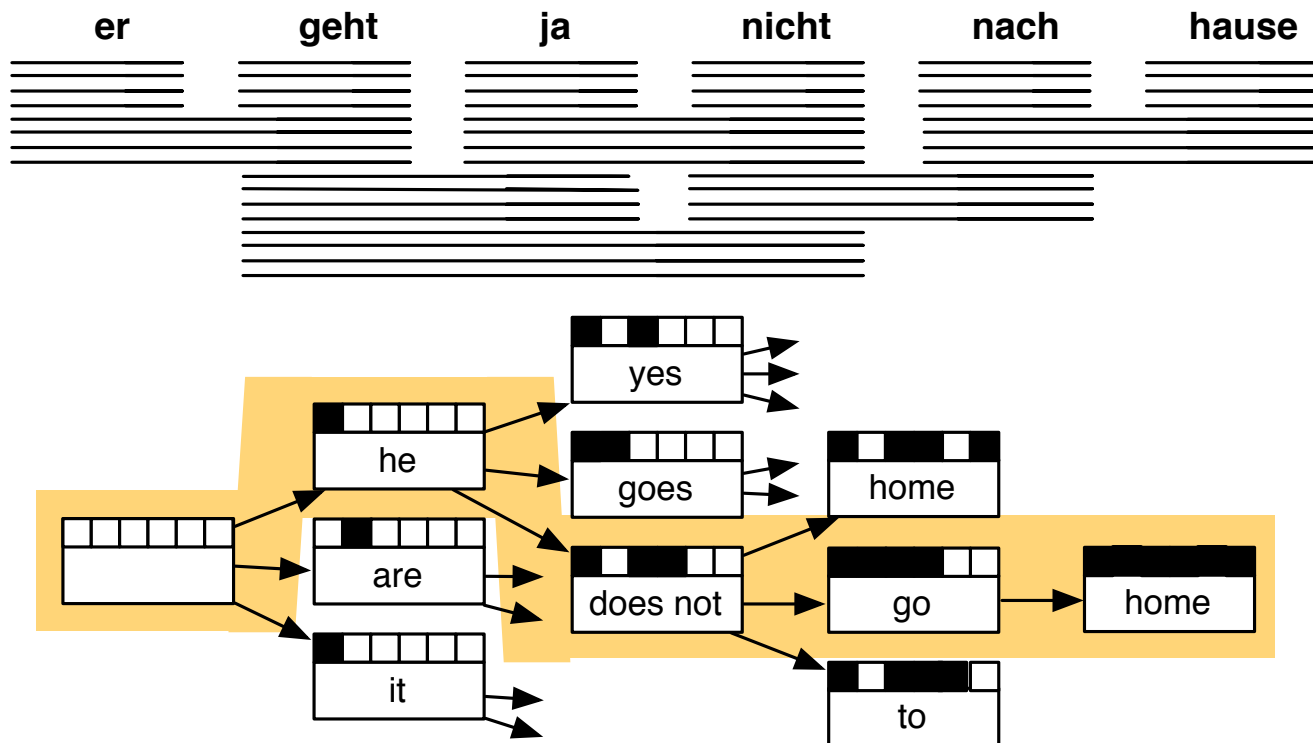
- ▶ After many steps, you get the large 'phrase table'. Each phrase in the source language can have many possible translations in the target language, and hence the search space can be combinatorially large!

## Translation Options

er	geht	ja	nicht	nach	hause
he	is	yes	not	after	house
it	are	is	do not	to	home
, it	goes	, of course	does not	according to	chamber
, he	go	,	is not	in	at home
it is		not		home	
he will be		is not		under house	
it goes		does not		return home	
he goes		do not		do not	
	is		to		
	are		following		
	is after all		not after		
	does		not to		
	not				
	is not				
	are not				
	is not a				

# Statistical Machine Translation

- ▶ Finally, you decode this hard search problem to find the best translation, e.g., using beam search on the several combinatorial paths through this phrase table (and also include the language model  $p(e)$  to rerank)



# Next Week

---

- ▶ IBM Alignment Model Details
- ▶ HMM Alignment Model
- ▶ Syntactic Models
- ▶ Neural Machine Translation (NMT)