Machine Translation Contd

Prof. Sameer Singh

CS 295: STATISTICAL NLP

WINTER 2017

March 7, 2017

Based on slides from Richard Socher, Chris Manning, Philipp Koehn, and everyone else they copied from.

Upcoming…

Project

- Status report due tonight: **March 7, 2017**
- Almost final report, only 5 pages

Summaries

- Paper summaries: March 14
- Summary 2 graded

Homework

- Homework 4 is due on March 13
- Write-up, code, and data released.
- **Lowest grade of the homeworks will be dropped**

Outline

Decoding Algorithms

Syntax-Based MT

Neural MT Models

Outline

Decoding Algorithms

Syntax-Based MT

Neural MT Models

Phrase Decoding: Stacks

function STACK DECODING(source sentence)

initialize stack with a null hypothesis
loop do
  pop best hypothesis h off of stack
  if h is a complete sentence, **return** h
  for each possible expansion h' of h
    assign a score to h'
    push h' onto stack

maximum size, B

h
E: english "Mary has"
F: * * _ _ _ _
c: partial score

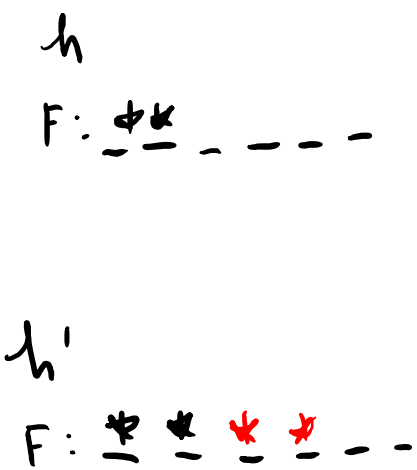
h'
E: "Mary has a"
F: * * * _ _ _
c': c + ___

h'
E: Mary has the
F: * * * _ _ _
c': c + ___

Monotonic Decoding

function STACK DECODING(source sentence)

initialize stack with a null hypothesis
loop do
  pop best hypothesis h off of stack
  if h is a complete sentence, **return** h
  for each possible expansion h' of h
    assign a score to h'
    push h' onto stack



Monotonic Phrase Decoding

Maria	no	dio	una	bofetada	a	la	bruja	verde
Mary	not	give	a	slap	to	the	witch	green
	did not		a slap		by		green witch	
	no		slap		to the			
	did not give				to			
					the			
		slap				the witch		

h

E: " "

F: _ _ _ _

c: 0.0

h'

E: "Mary"

F: _ _ _ _

c: $\log p(\text{Mary} \mid \text{Maria}) B = 1$

$+ \log(\text{Mary} \mid \langle s \rangle, \langle s \rangle)$

Monotonic Phrase Decoding

Maria	no	dio	una	bofetada	a	la	bruja	verde
Mary	not	give	a	slap	to	the	witch	green
	did not		a slap		by		green witch	
	no	slap			to the			
	did not give				to			
					the			
		slap				the witch		

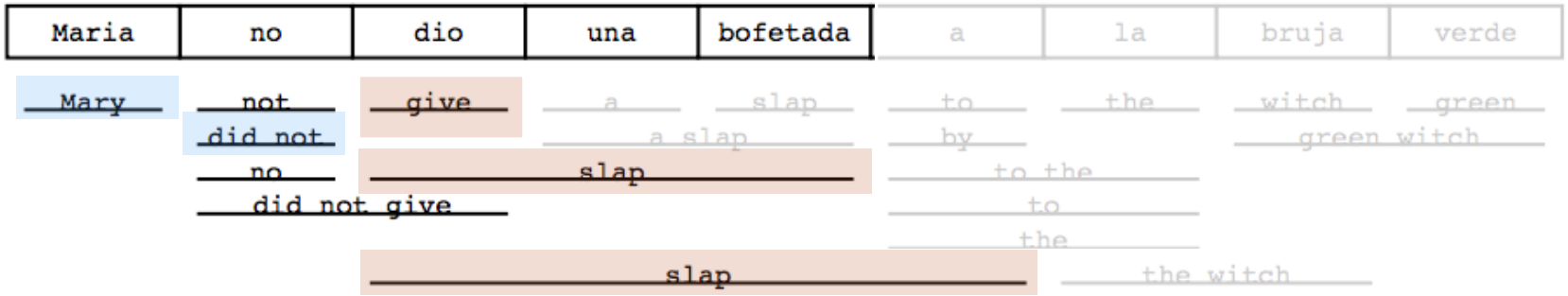
(Mary) → "not"

(Mary) → "did not"

(Mary) → "no"

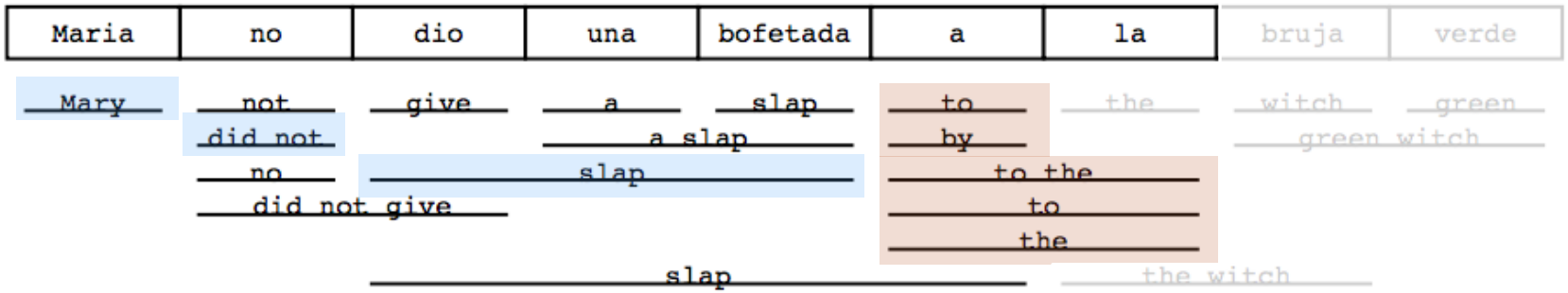
(Mary) → "did not give"

Monotonic Phrase Decoding



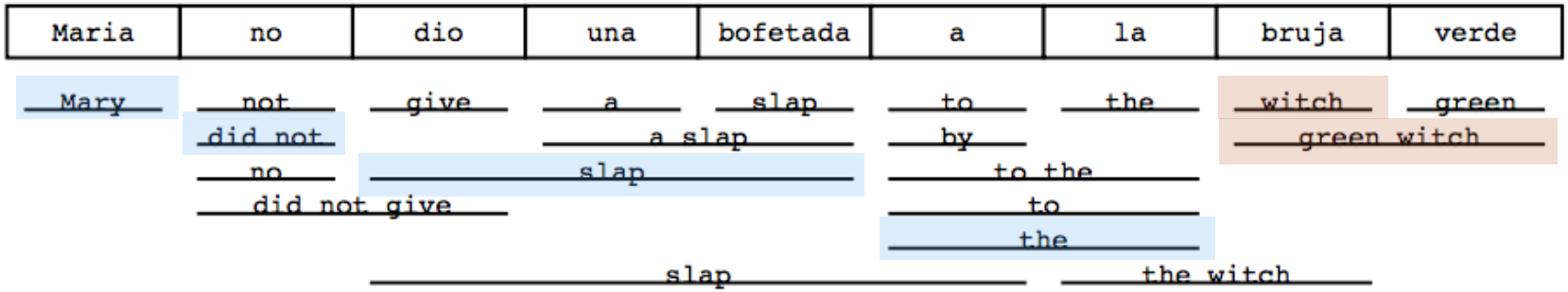
(Mary) (did not)

Monotonic Phrase Decoding



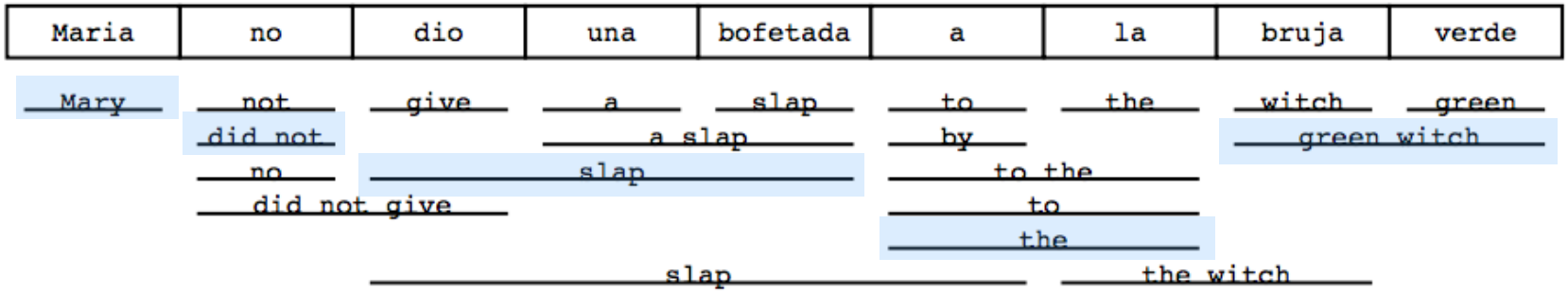
(Mary) (did not) (slap)

Monotonic Phrase Decoding



(Mary) (did not) (slap) (the)

Monotonic Phrase Decoding



(Mary) (did not) (slap) (the) (green witch)

Non-Monotonic Decoding

h E: P_1 + P_2

F: * * _ * *

C:

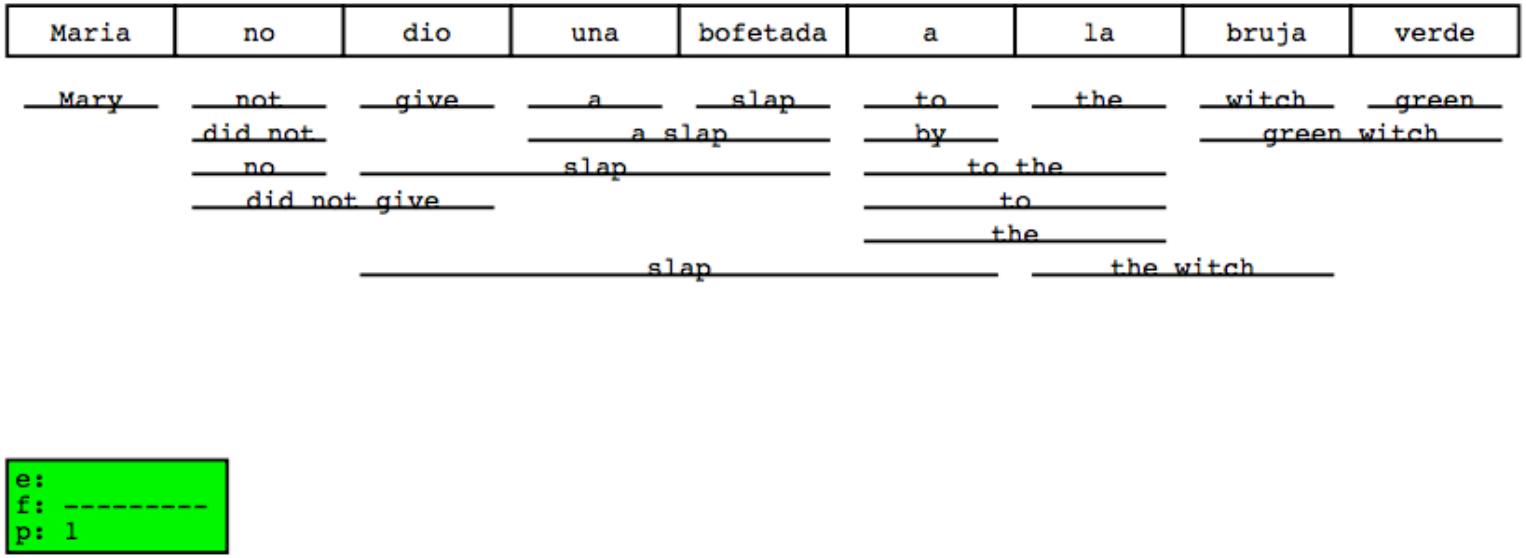
h'

function STACK DECODING(source sentence)

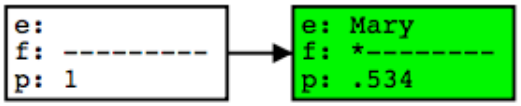
initialize stack with a null hypothesis
loop do
  pop best hypothesis h off of stack
  if h is a complete sentence, **return** h
  for each possible expansion h' of h
    assign a score to h'
    push h' onto stack

any P that doesn't overlap w/ h.F

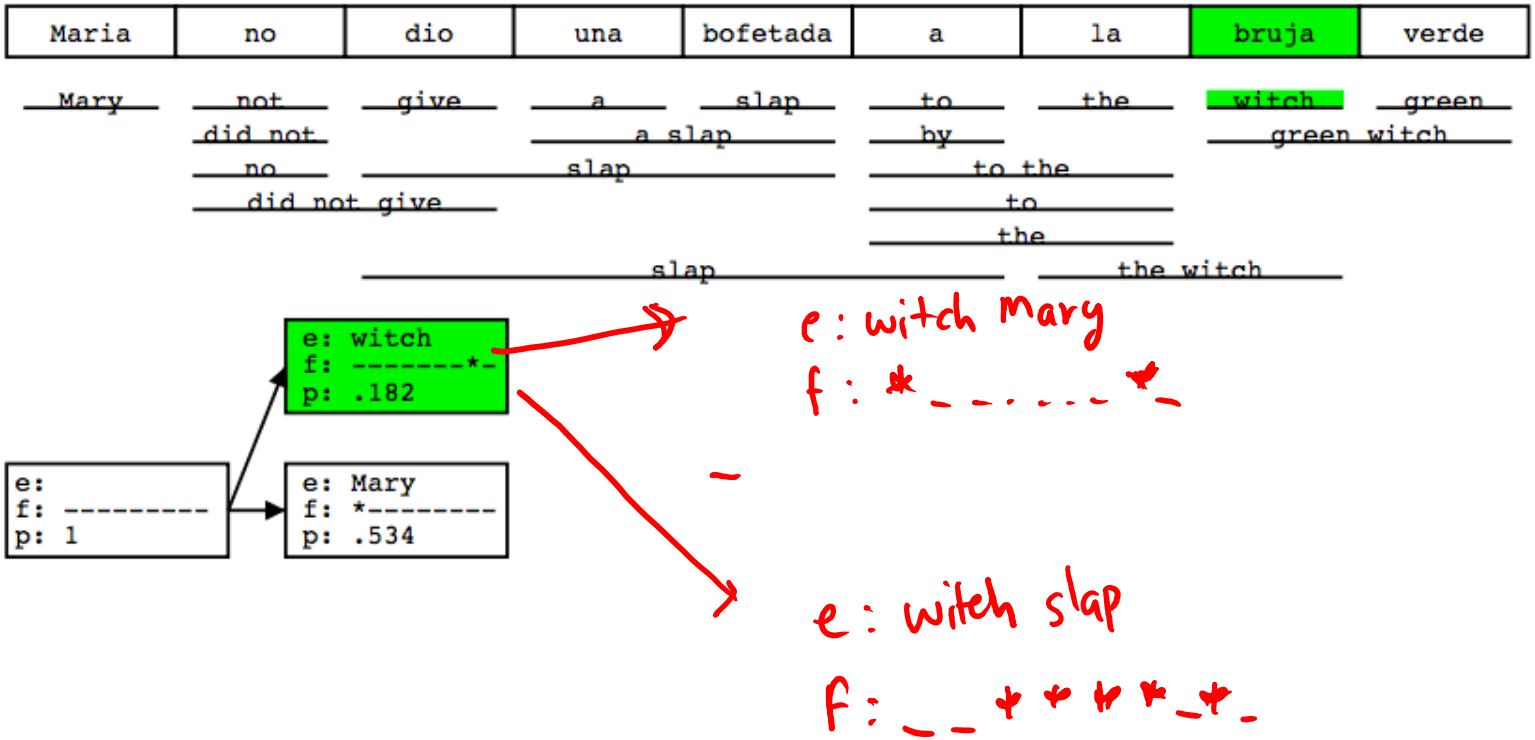
Non-Monotonic Decoding



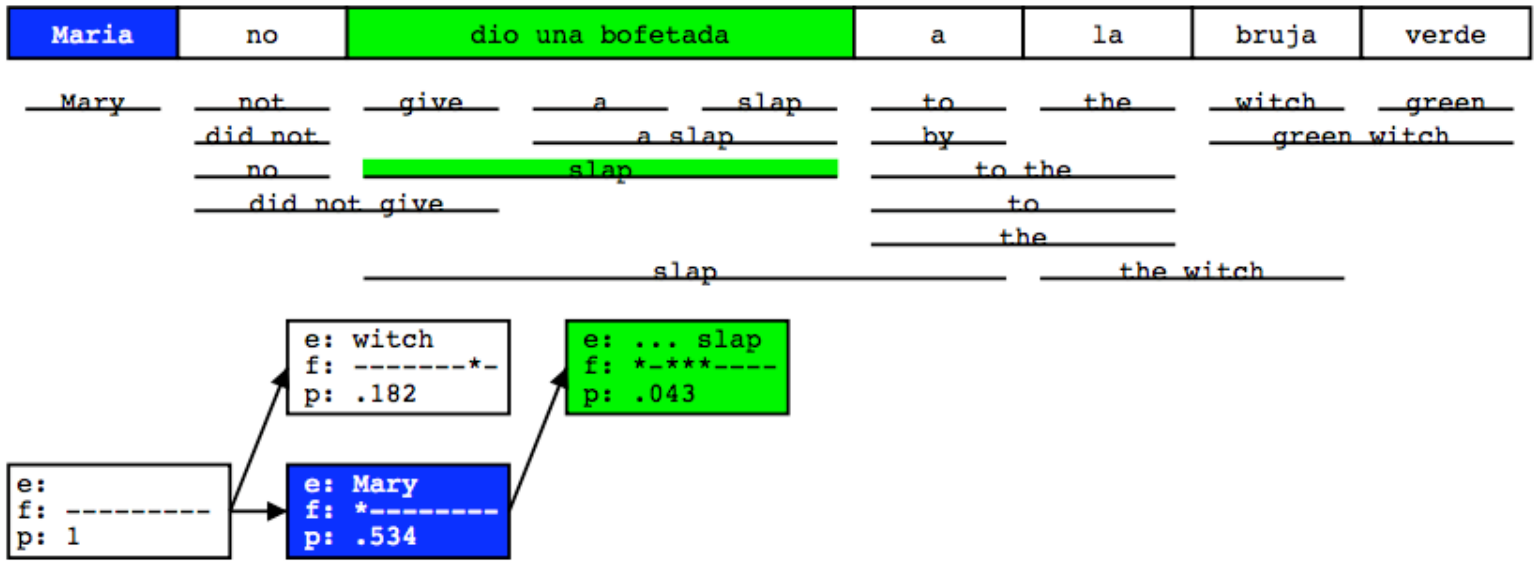
Non-Monotonic Decoding



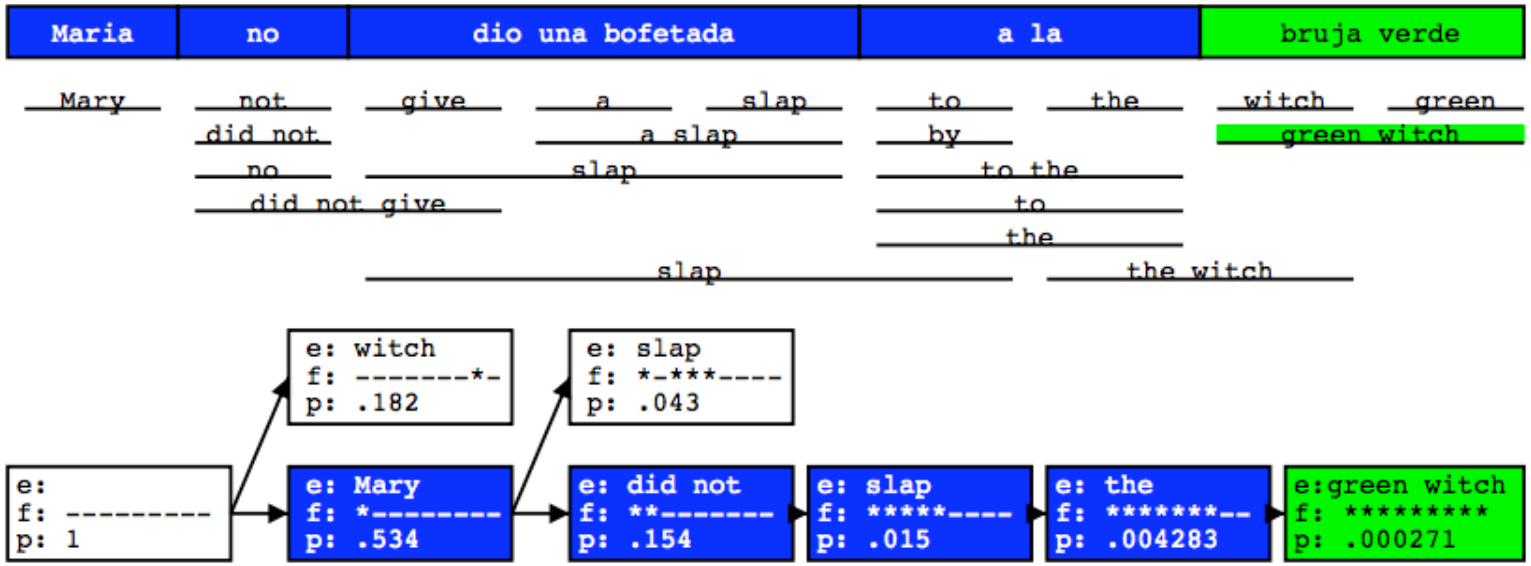
Non-Monotonic Decoding



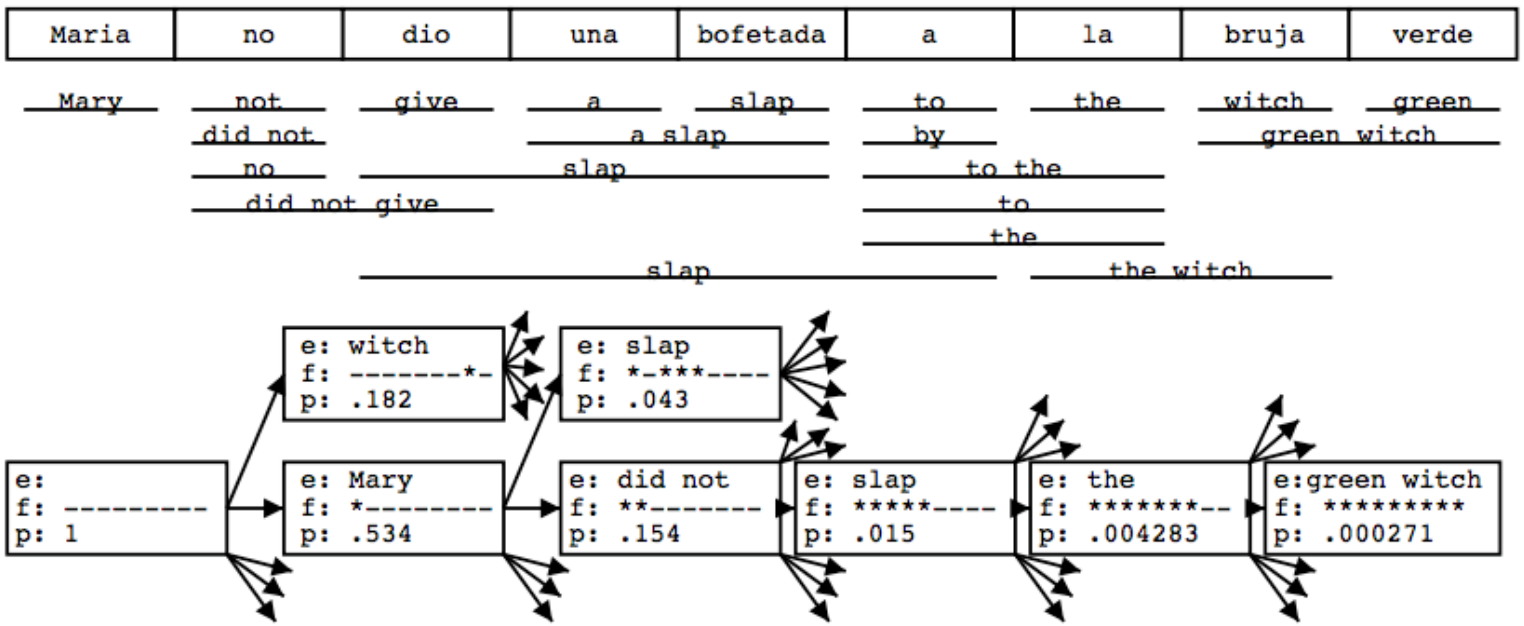
Non-Monotonic Decoding



Non-Monotonic Decoding



Non-Monotonic Decoding



Comparing Partial Translations

h_1 E: "Mary"

F: * _ _ _ _

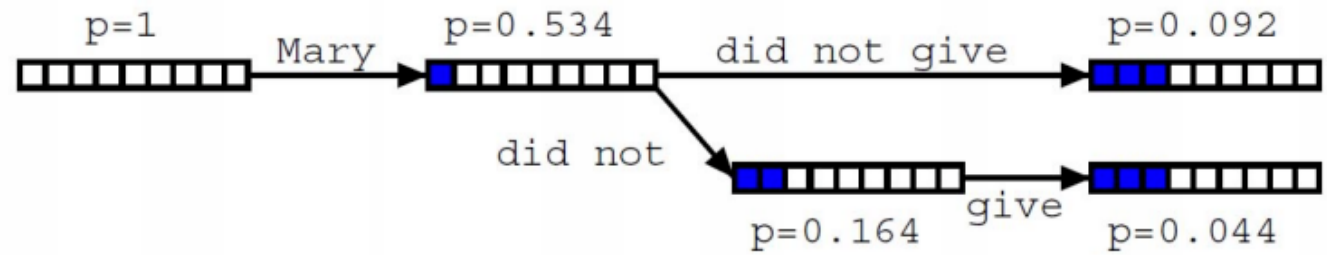
c:

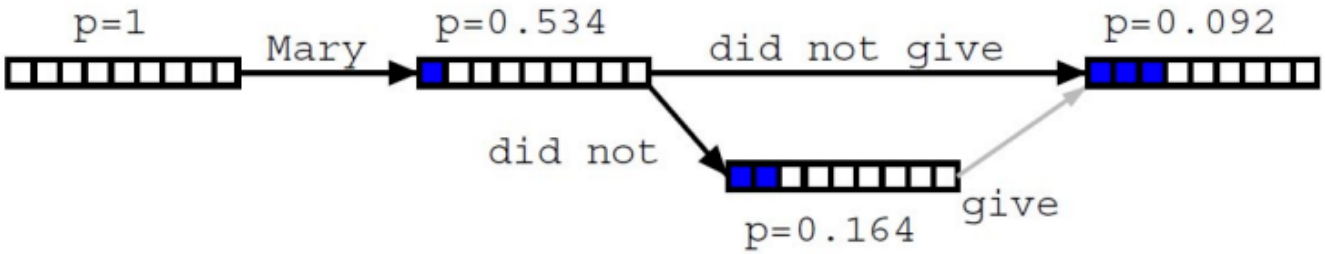
h_2 E: "Mary did not slap the"

F: * * * * * * _ _

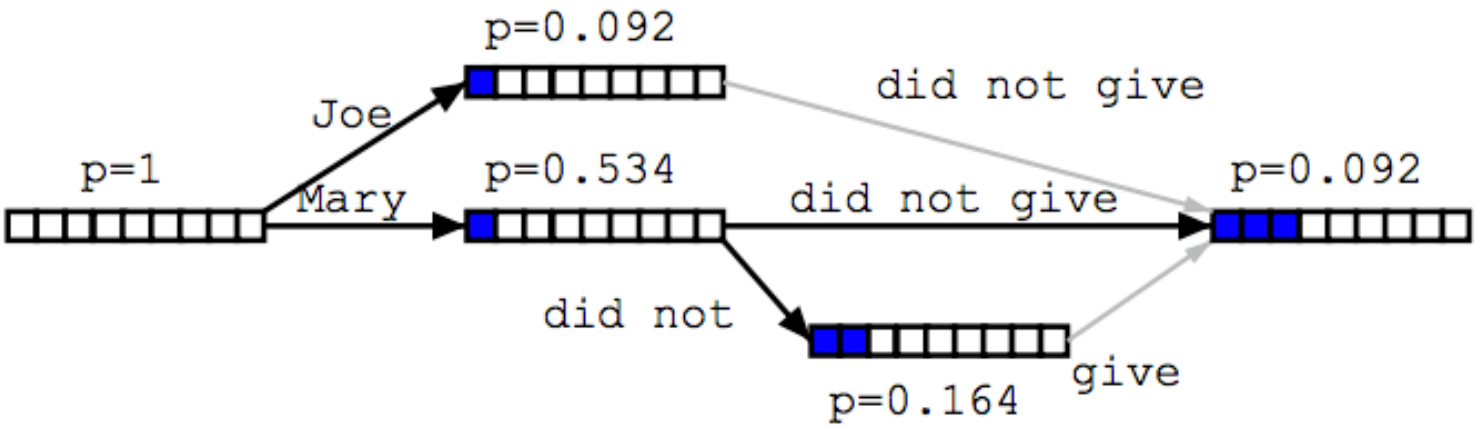
c:

Hypothesis Recombination





Hypothesis Recombination



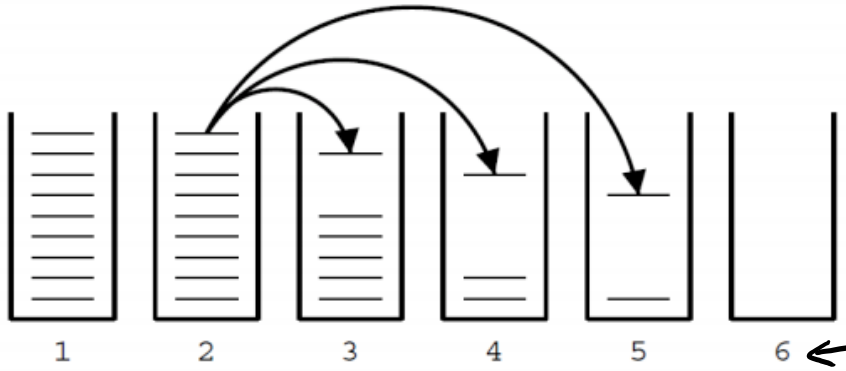
Multi-Stack Decoding

function STACK DECODING(source sentence)

initialize stack with a null hypothesis
loop do
  pop best hypothesis h off of stack
  if h is a complete sentence, **return** h
  for each possible expansion h' of h
    assign a score to h'
    push h' onto stack

function STACKDECODING(source sentence)

initialize stacks from 0..n
for i = 0 .. n-1
  for *h* in k-best hypothesis from stack[*i*]
  for *h'* in possible expansion of *h*
    assign a score to *h'*
    push *h'* onto stack[len(*h'*)]
return max(stack[*n*])



|h.F|
number of 1s in |h.F|

Outline

Decoding Algorithms

Syntax-Based MT

Neural MT Models

Limits of the Phrase Model

Non-Contiguous Phrases

Ich habe das Auto gekauft

I bought the car

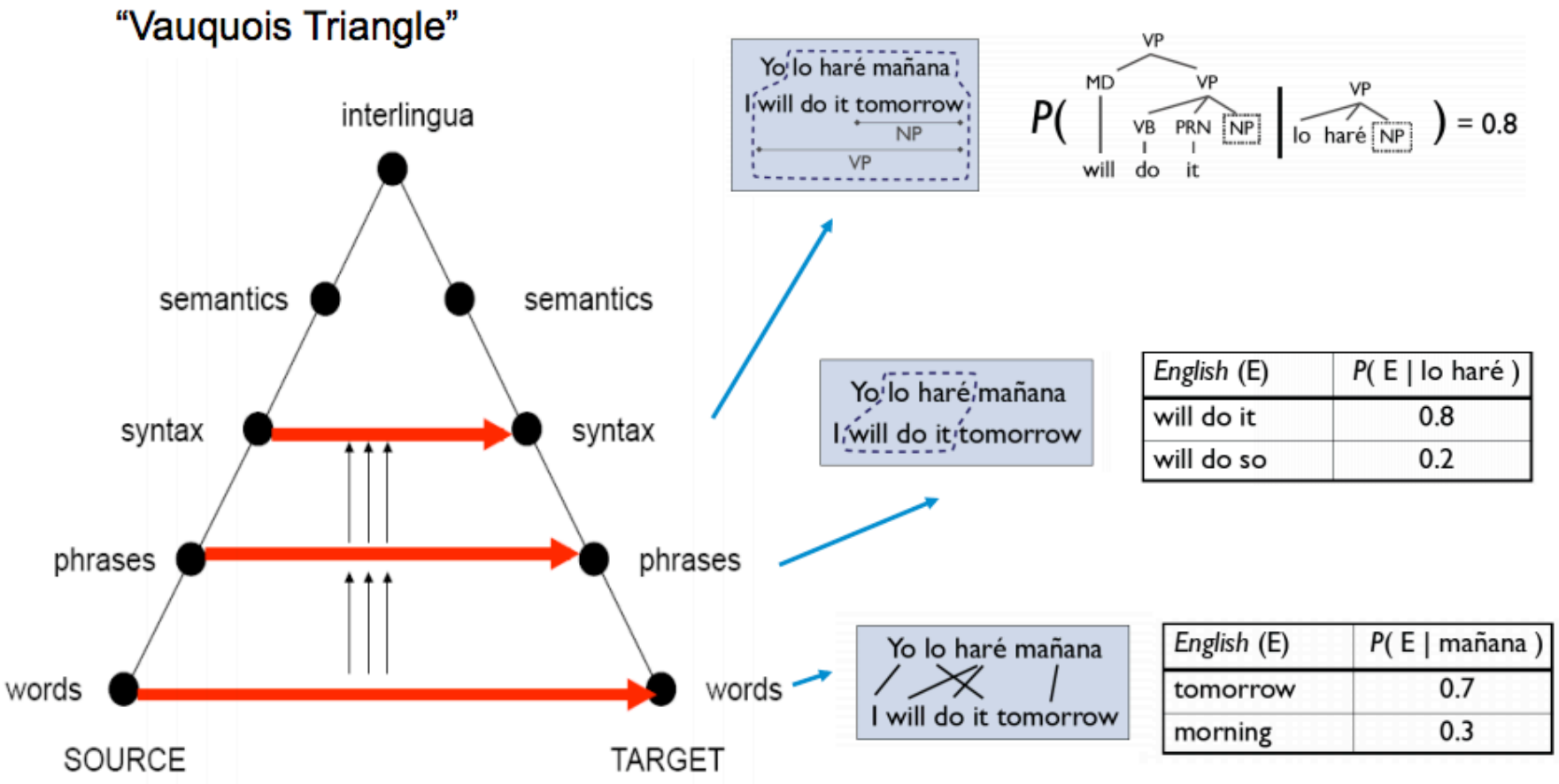
Syntactic Transformations

Den Antrag verabschiedet das Parlament

The draft approves the Parliament

Syntax leads to Good English

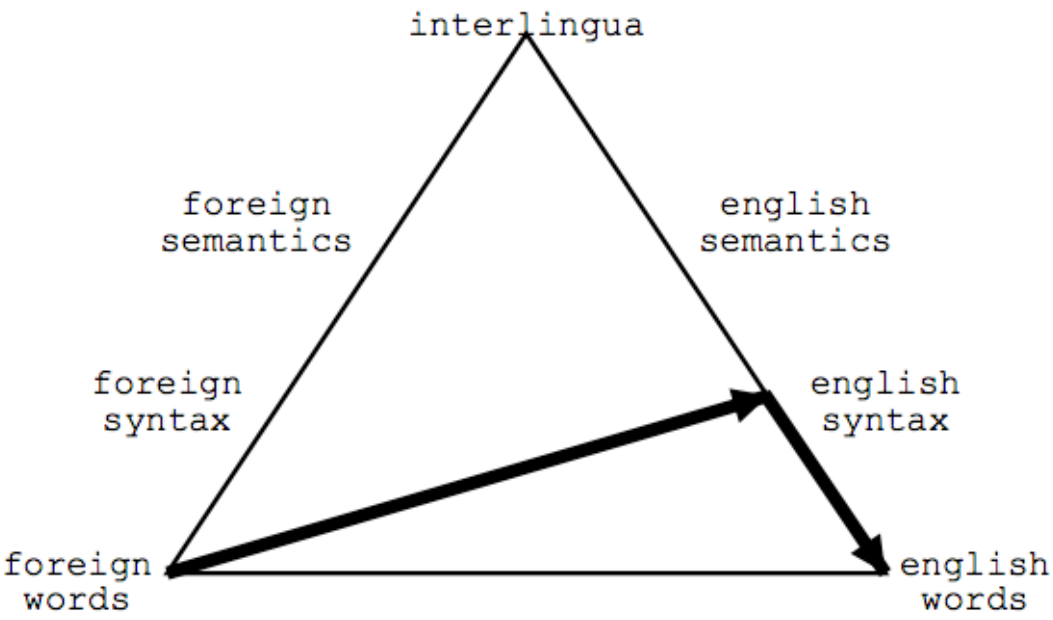
The Vauquios Triangle



English (E)	*P*(E | lo haré)
will do it	0.8
will do so	0.2

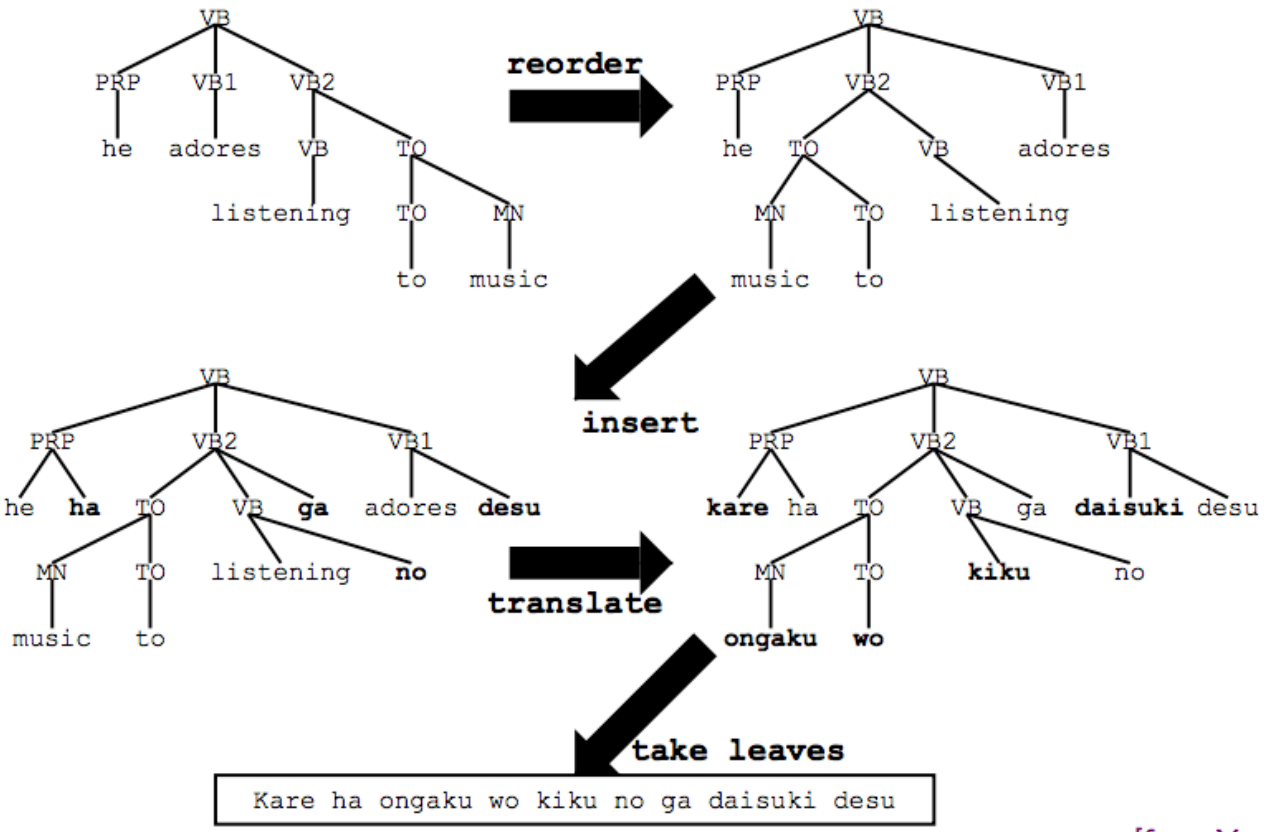
English (E)	*P*(E | mañana)
tomorrow	0.7
morning	0.3

String to Tree Translation



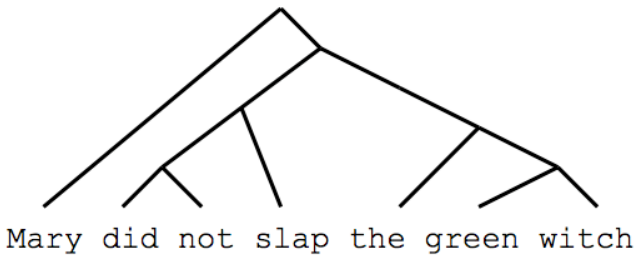
[Yamada and Knight, 2001]

String to Tree Translation

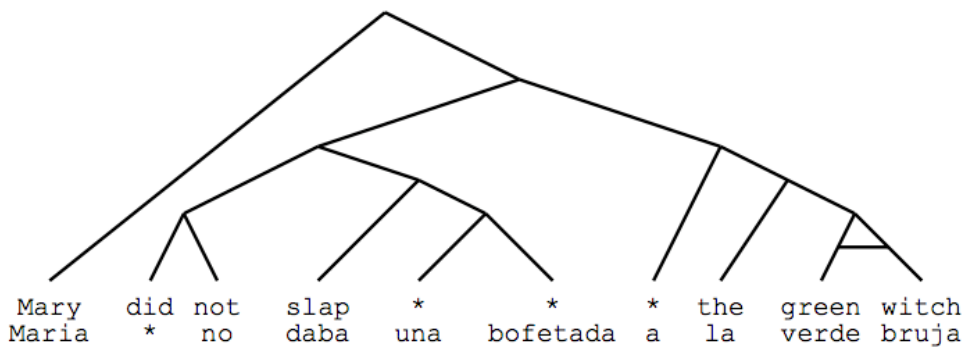


[from Yamada and Knight, 2001]

Synchronous CFGs



$$A \rightarrow A_1 A_2 \| A_1 A_2$$
$$A \rightarrow A_1 A_2 \| A_2 A_1$$
$$A \rightarrow e \| f$$
$$A \rightarrow e \| *$$
$$A \rightarrow * \| f$$



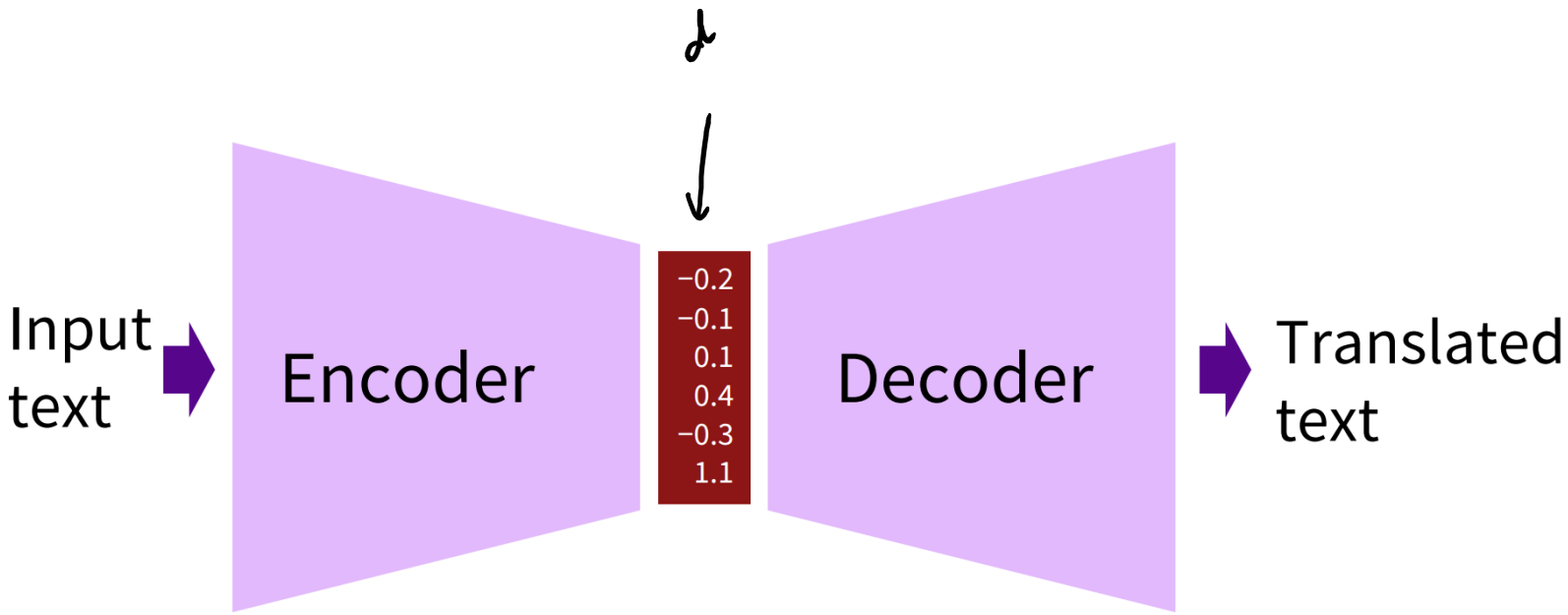
Outline

Decoding Algorithms

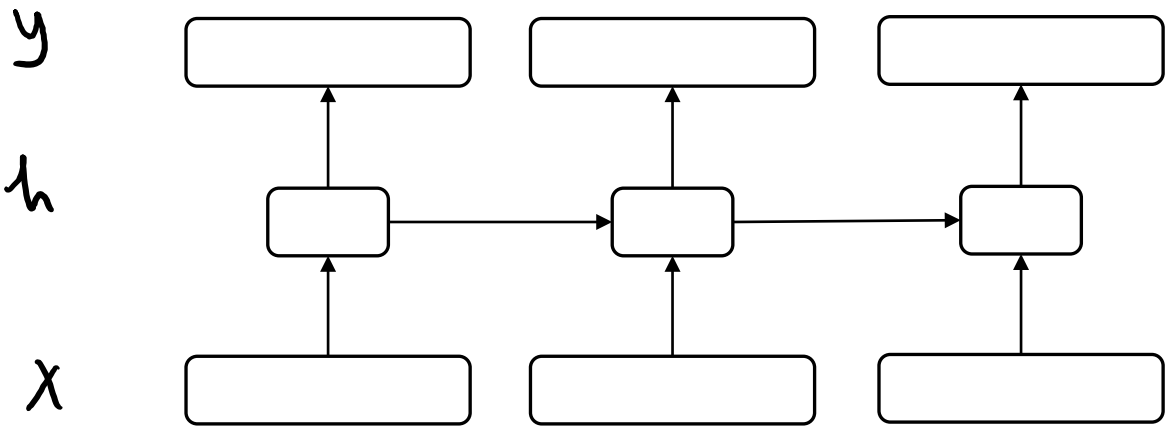
Syntax-Based MT

Neural MT Models

Neural MT Models



Recurrent Neural Networks

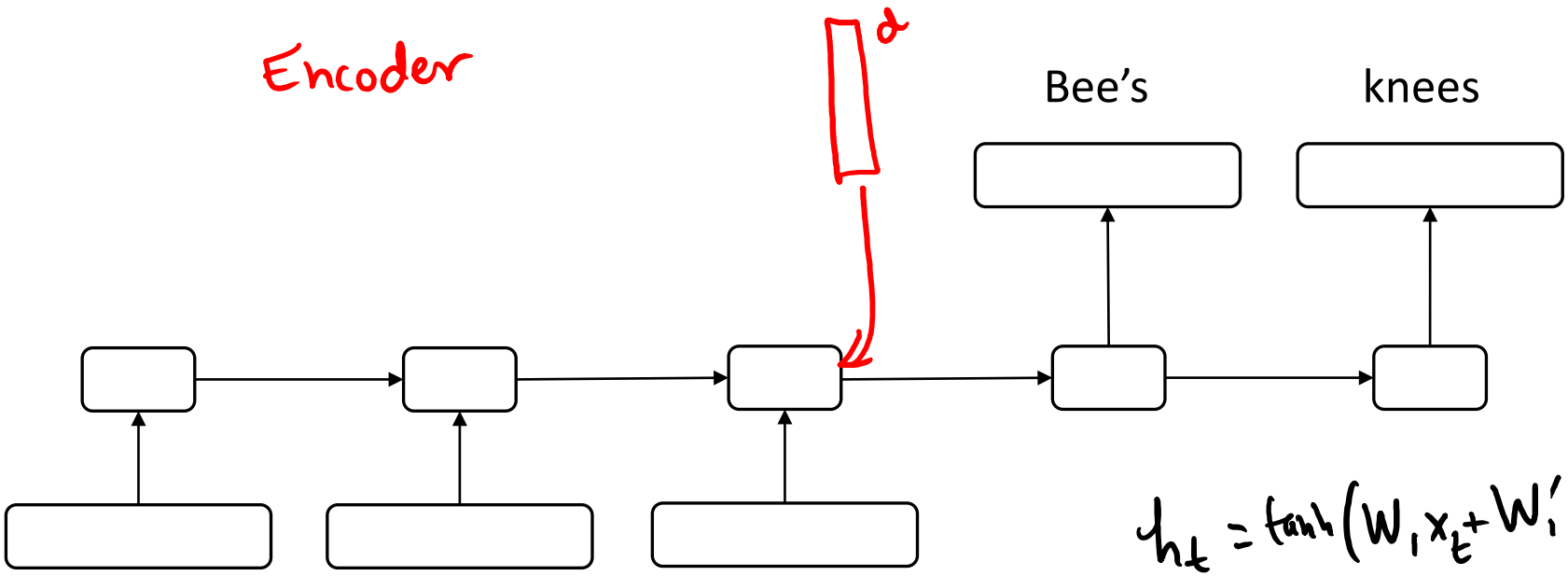


$y_t = \text{softmax}(W_0 h_t)$

$h_t = (W_1 x_t + W_1' h_{t-1})$

tanh

Recurrent Neural Networks



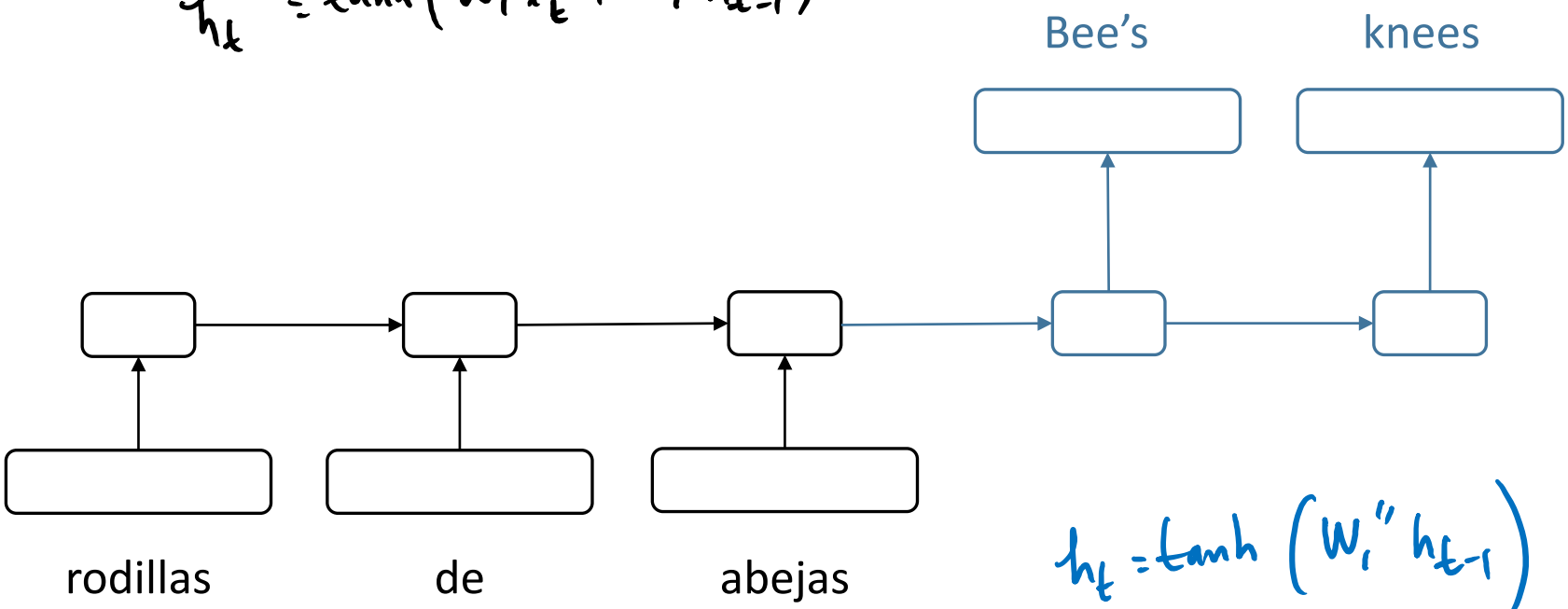
$$h_t = \tanh(W_1 x_t + W_1' h_{t-1})$$

$$y_t = \sigma(W_0 h_t)$$

Decoder

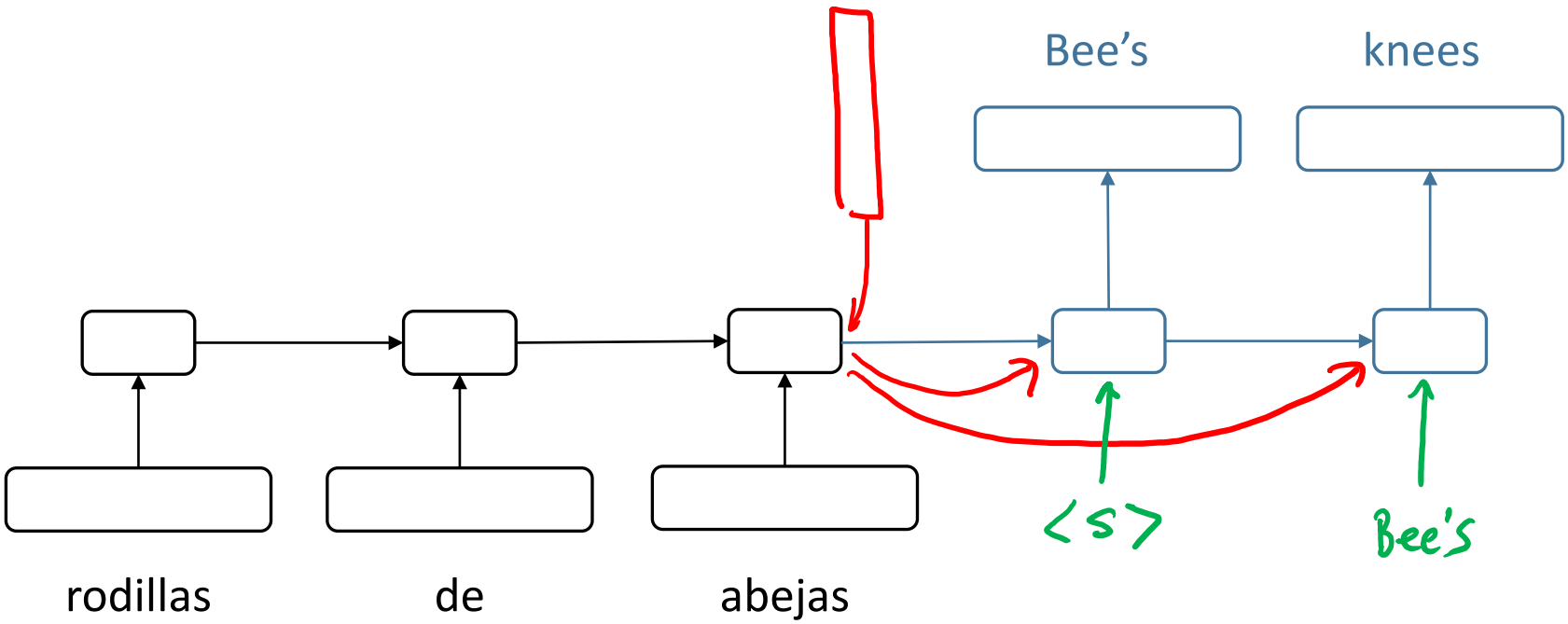
Different Weights

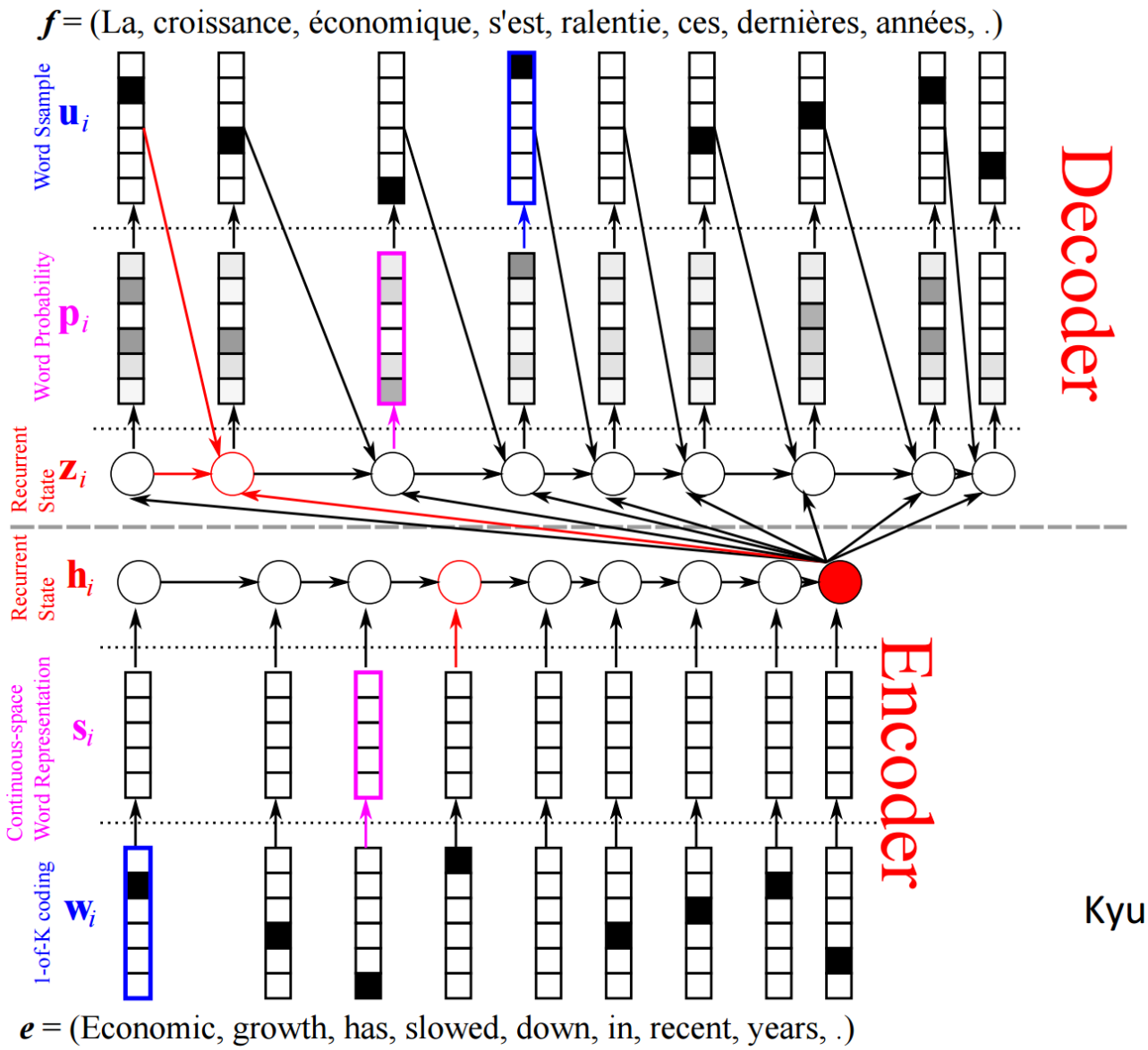
$h_t = \tanh(W_1 x_t + W_1' h_{t-1})$



$h_t = \tanh(W_1'' h_{t-1})$

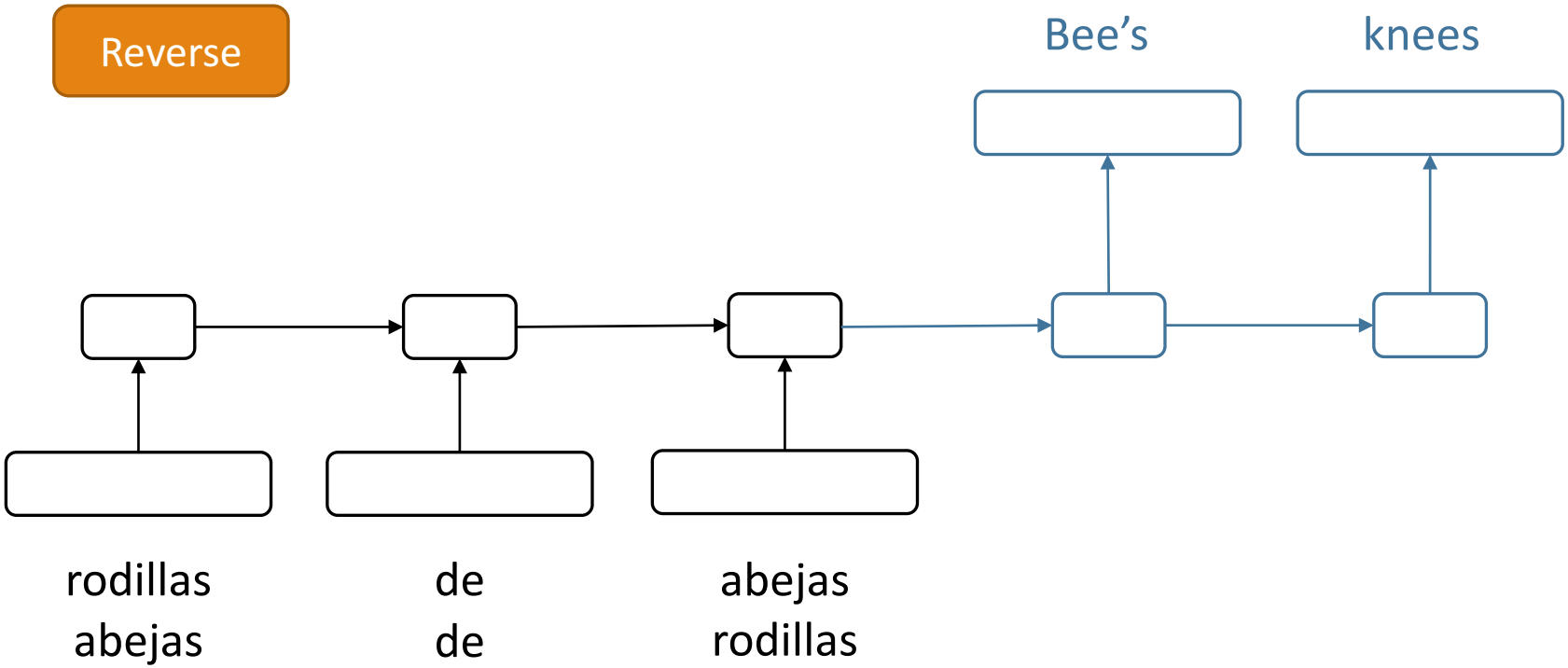
More Connections



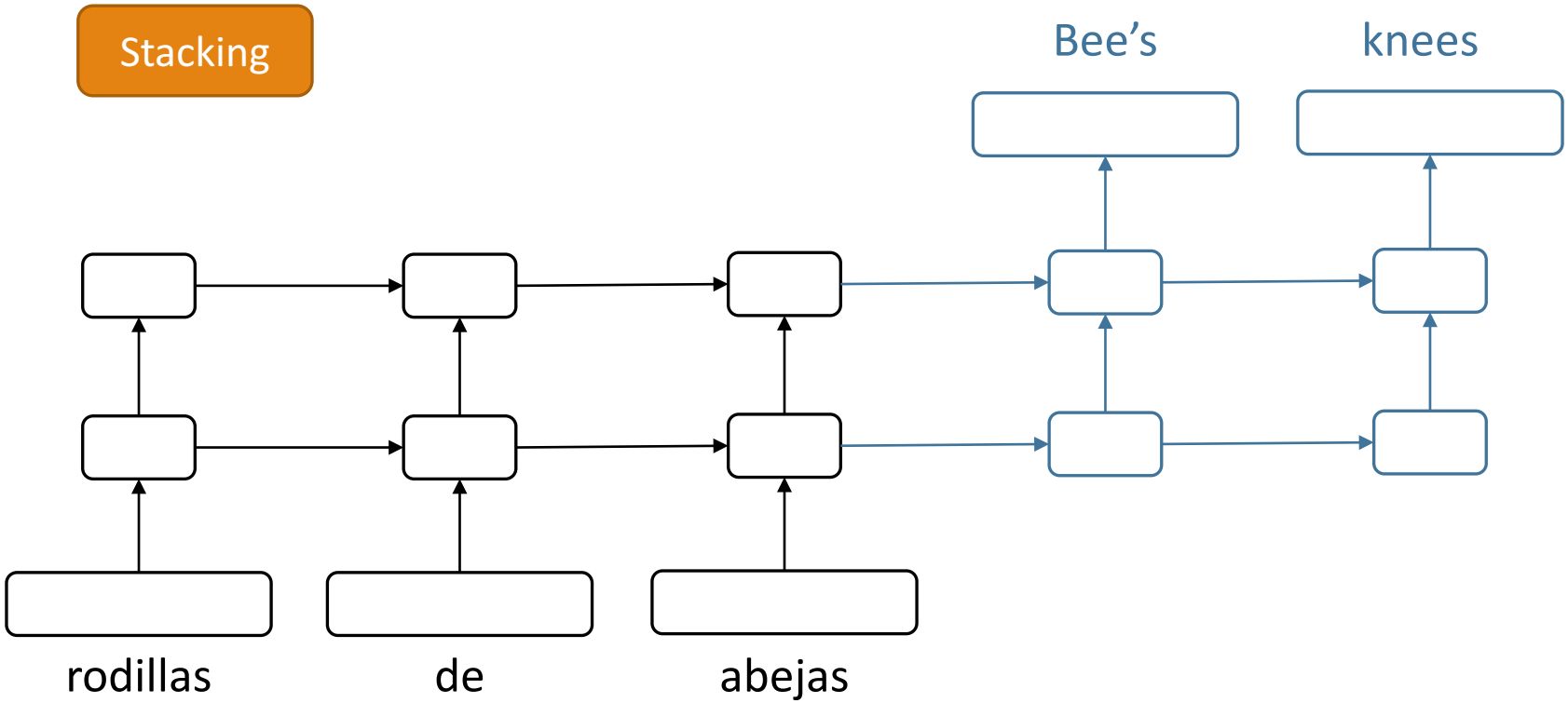


Kyunghyun Cho et al. 2014

Other Extensions

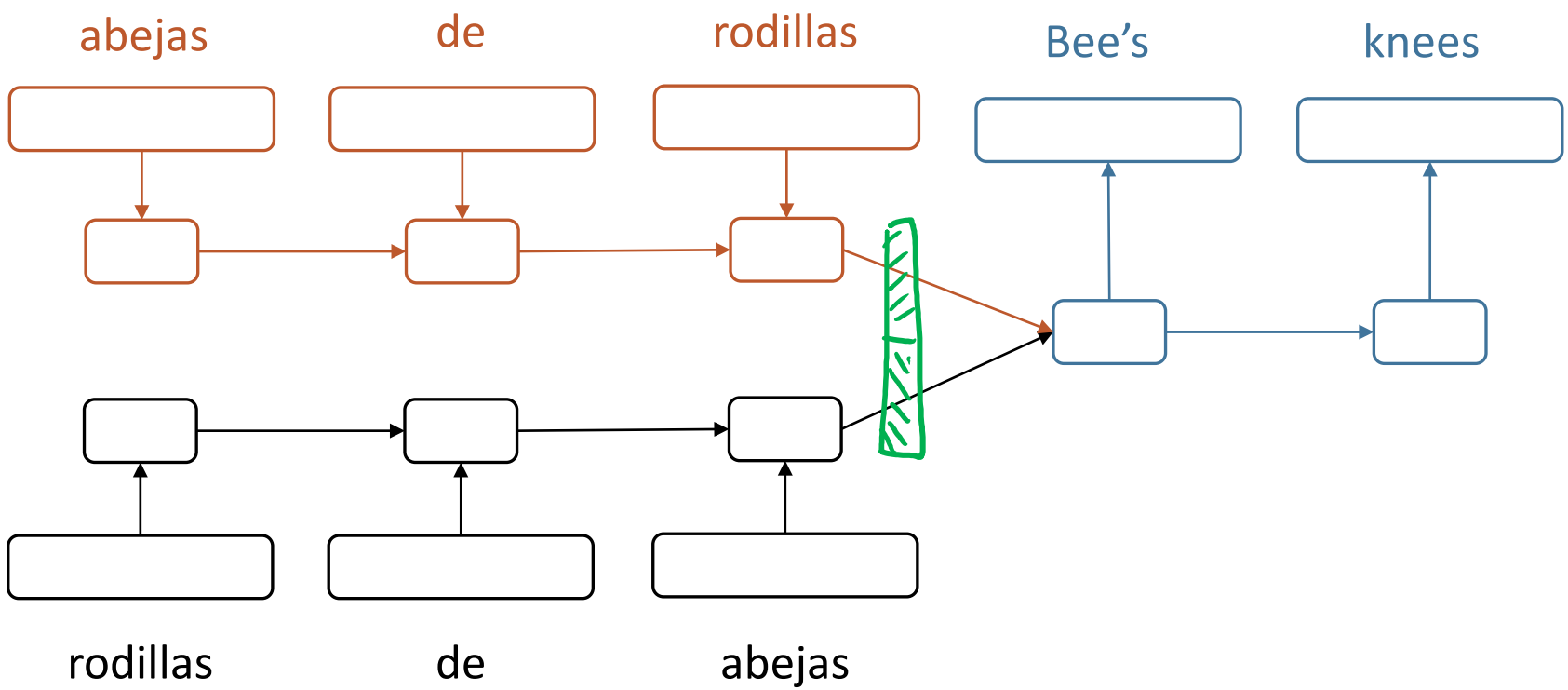


Other Extensions

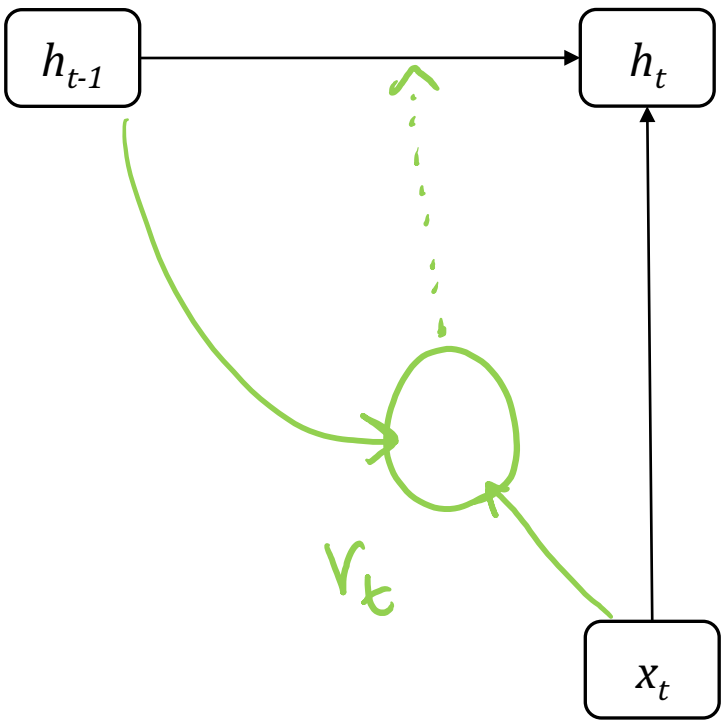


Other Extensions

Bi-directional



Regular Recurrent Units

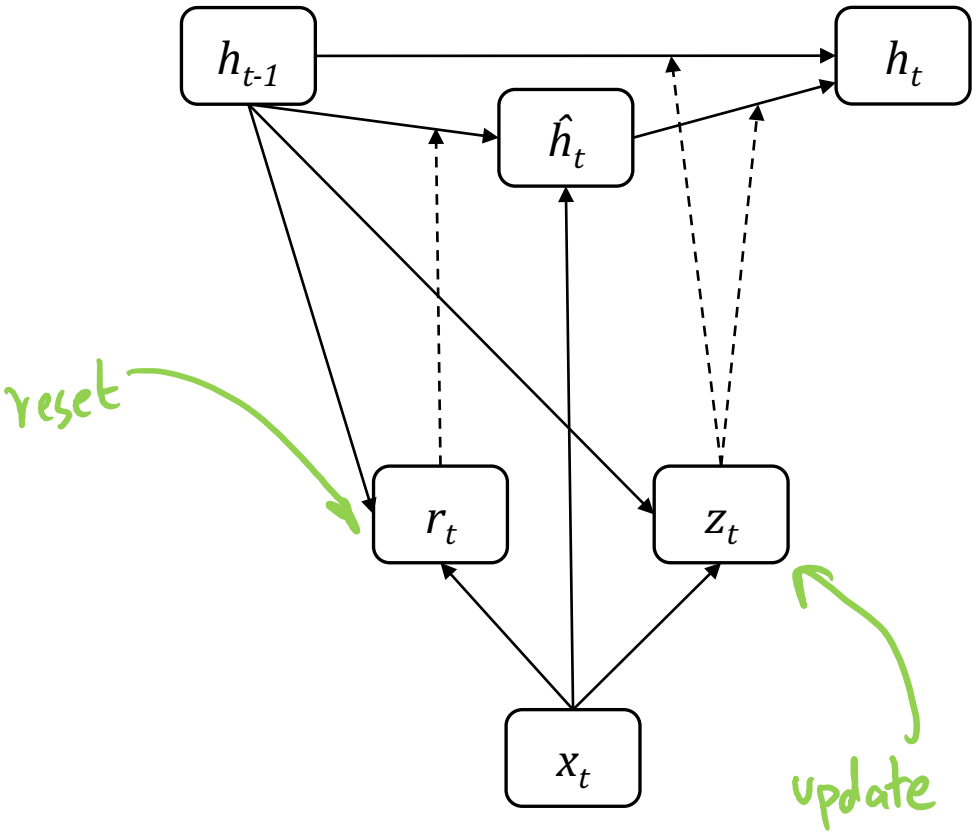


$$h_t = \tanh\left(W_1 x_t + W_1' h_{t-1}\right)$$

$$h_t = \tanh\left(W_1 x_t + \overset{W_1'}{r_t} \circ h_{t-1}\right)$$

$$r_t = \text{sigmoid}\left(W_2 x_t + W_2' h_{t-1}\right)$$

Gated Recurrent Units



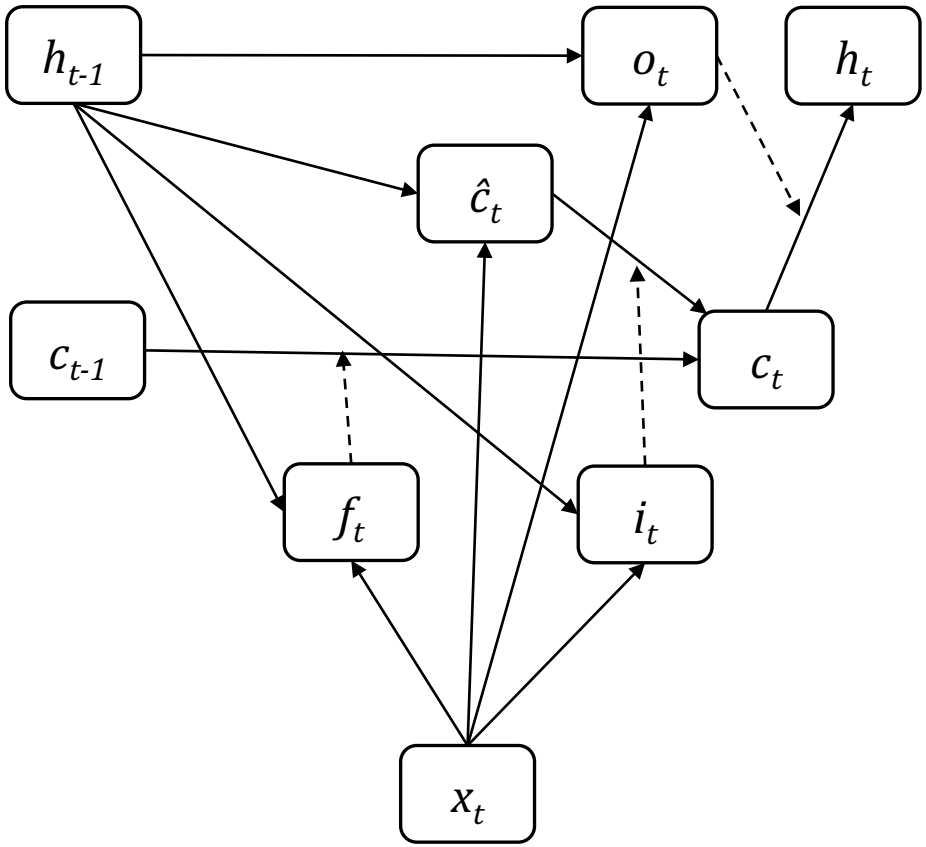
$$z_t = \sigma\left(W^{(z)}x_t + U^{(z)}h_{t-1}\right)$$

$$r_t = \sigma\left(W^{(r)}x_t + U^{(r)}h_{t-1}\right)$$

$$\tilde{h}_t = \tanh\left(Wx_t + r_t \circ Uh_{t-1}\right)$$

$$h_t = z_t \circ h_{t-1} + (1 - z_t) \circ \tilde{h}_t$$

Long Short-Term Memory



$$i_t = \sigma\left(W^{(i)}x_t + U^{(i)}h_{t-1}\right)$$

$$f_t = \sigma\left(W^{(f)}x_t + U^{(f)}h_{t-1}\right)$$

$$o_t = \sigma\left(W^{(o)}x_t + U^{(o)}h_{t-1}\right)$$

$$\tilde{c}_t = \tanh\left(W^{(c)}x_t + U^{(c)}h_{t-1}\right)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t$$

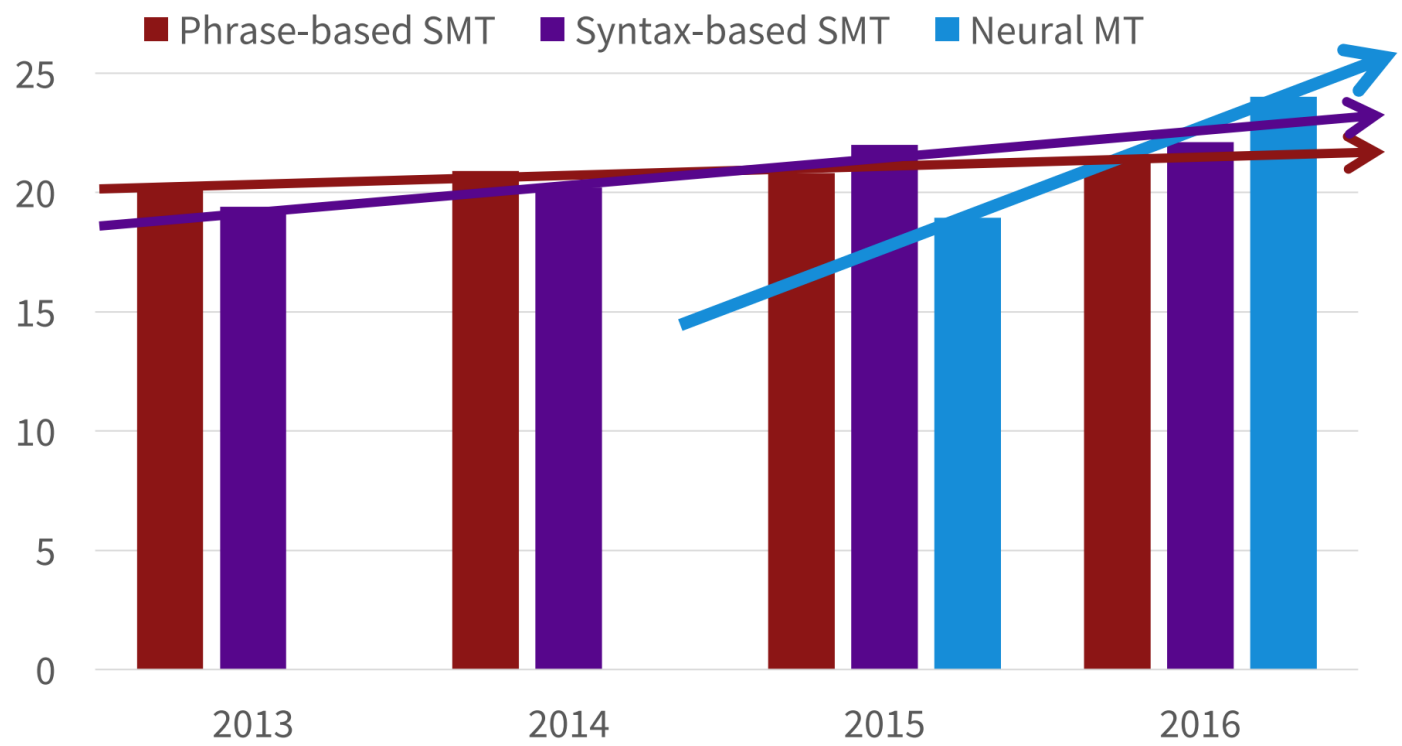
$$h_t = o_t \circ \tanh(c_t)$$

Neural MT Results

Method	test BLEU score (ntst14)
Baseline System [29]	33.30
Cho et al. [5]	34.54
Best WMT'14 result [9]	**37.0**
Rescoring the baseline 1000-best with a single forward LSTM	35.61
Rescoring the baseline 1000-best with a single reversed LSTM	35.85
Rescoring the baseline 1000-best with an ensemble of 5 reversed LSTMs	**36.5**
Oracle Rescoring of the Baseline 1000-best lists	~45

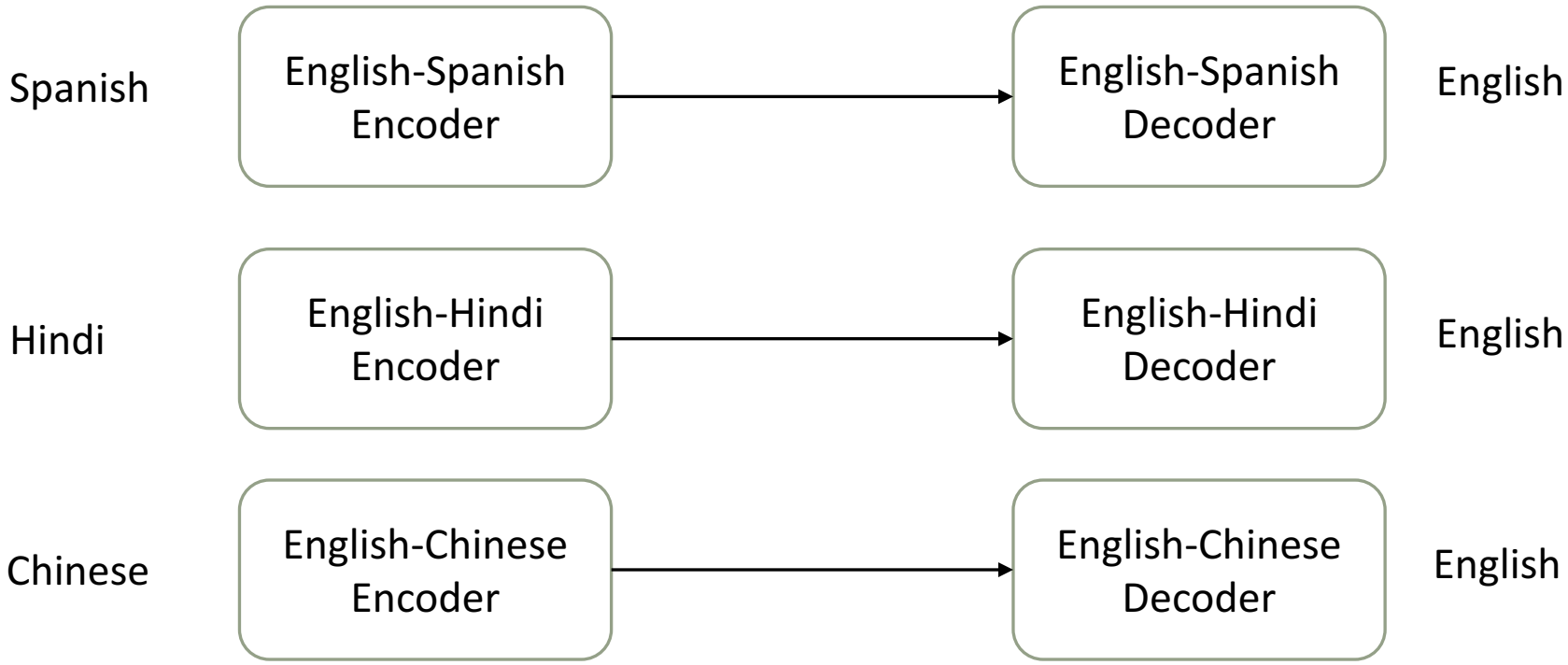
Sequence to Sequence Learning by Sutskever et al. 2014

Trend in Machine Translation

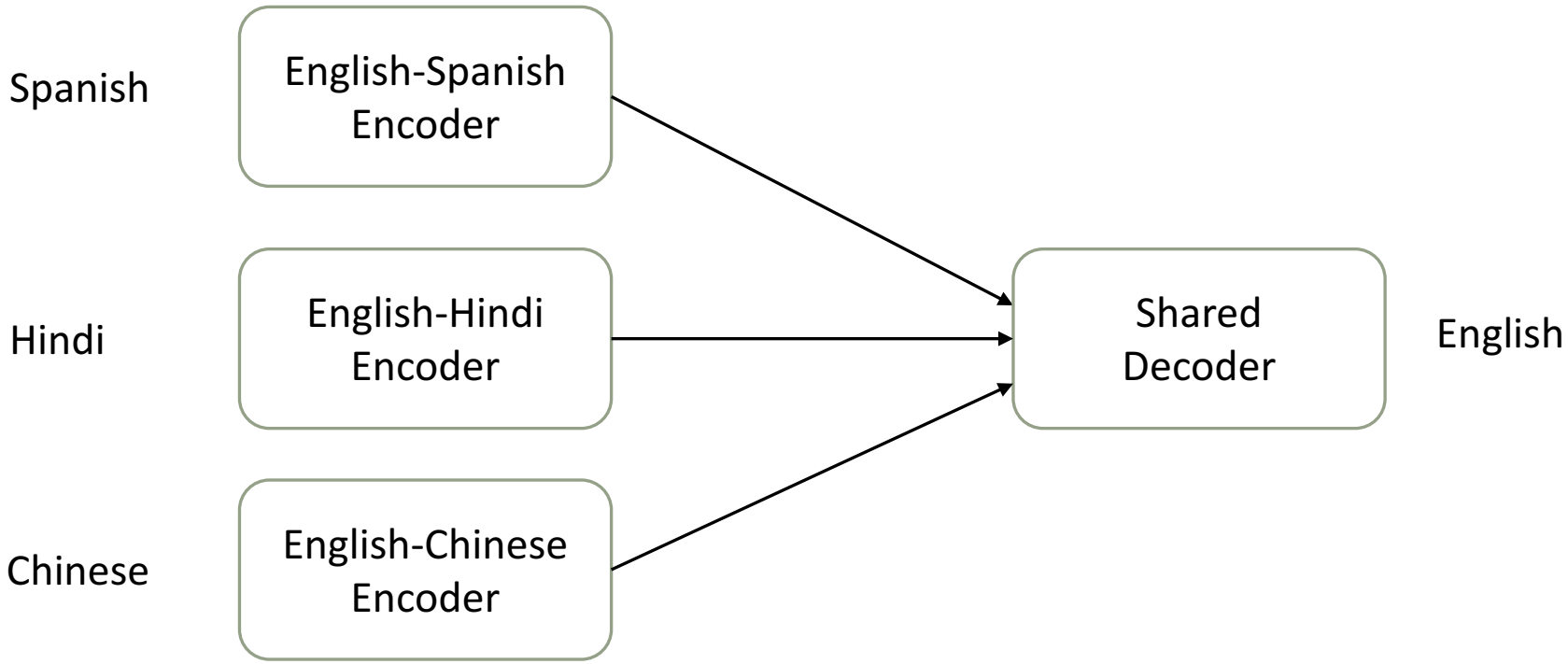


From [Sennrich 2016, http://www.meta-net.eu/events/meta-forum-2016/slides/09_sennrich.pdf]

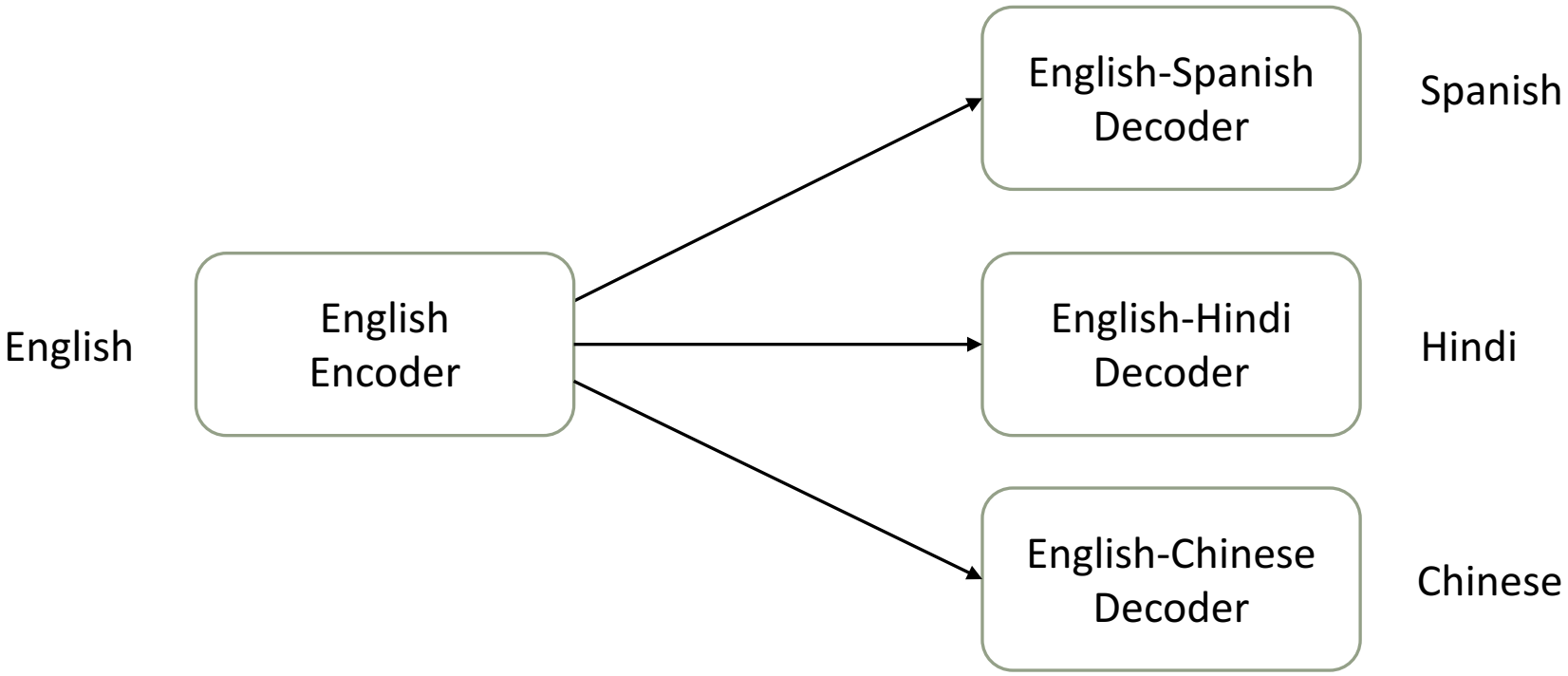
Multilingual Neural MT: Naïve



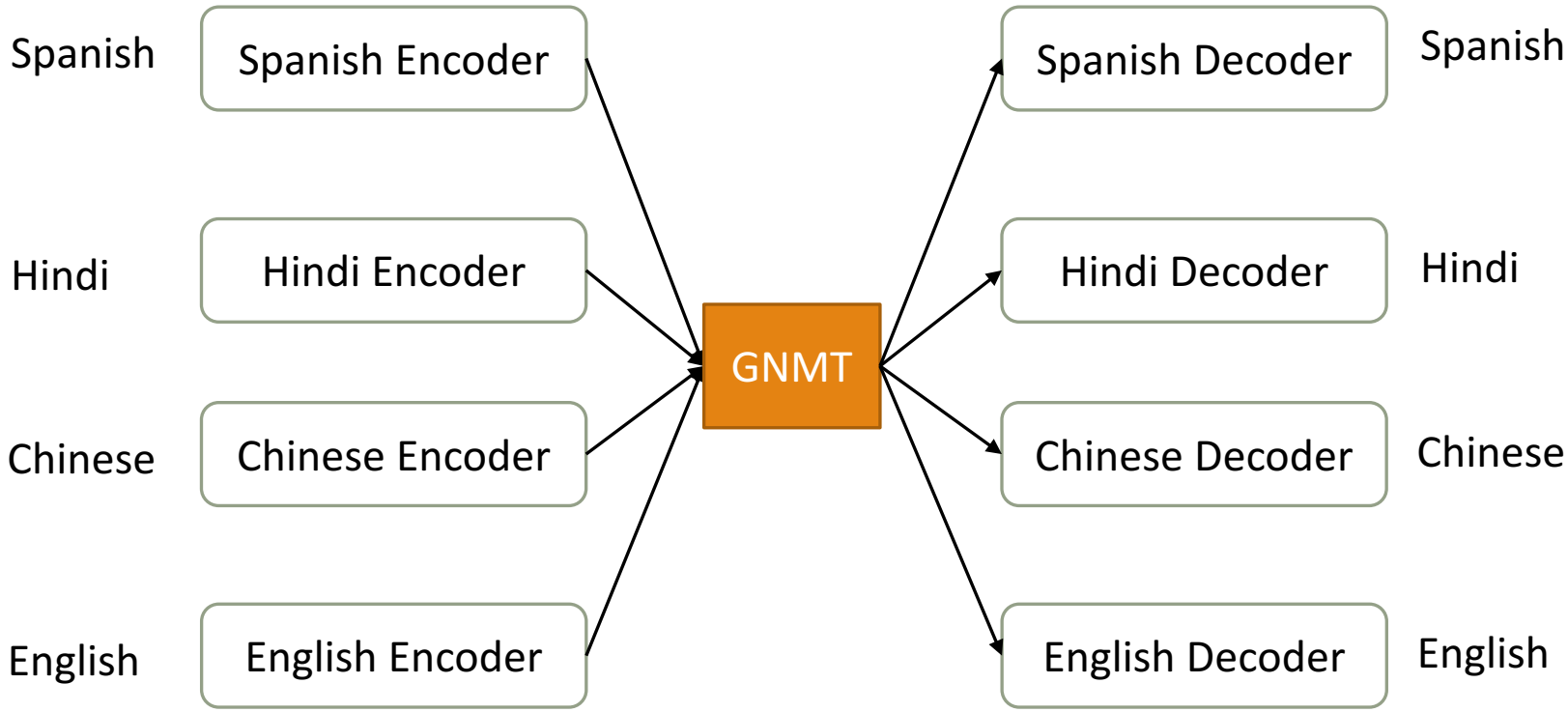
Multilingual NMT: Decoder



Multilingual NMT: Encoder

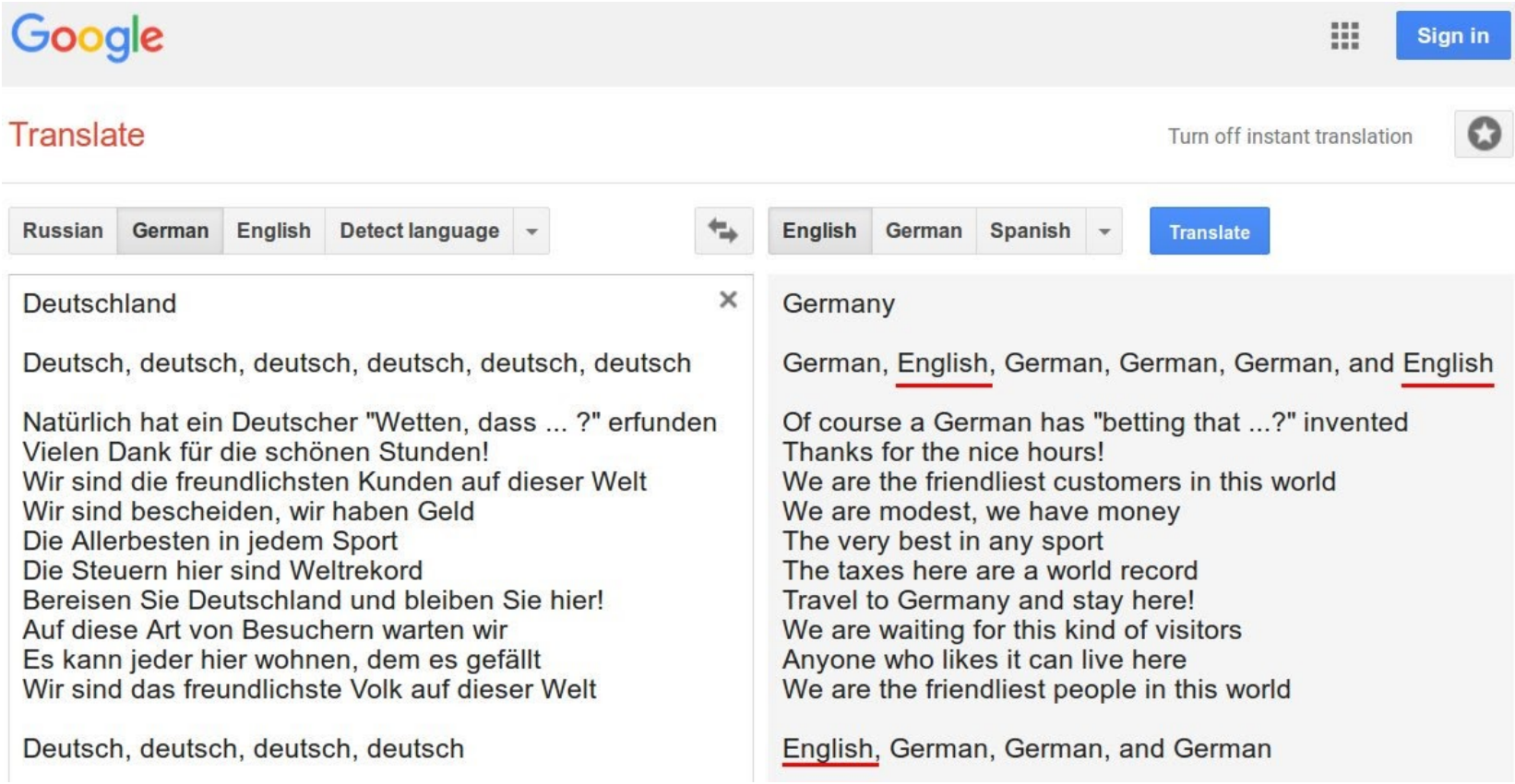


Google Neural MT



Out of Control!

Out of Control!



Out of Control!