

Machine Translation Contd

Prof. Sameer Singh

CS 295: STATISTICAL NLP

WINTER 2017

March 7, 2017

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

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

h

F: ~~⊗~~ _ _ _ _

h'

F: ~~⊗~~ ~~⊗~~ ~~⊗~~ ~~⊗~~ _ _

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		

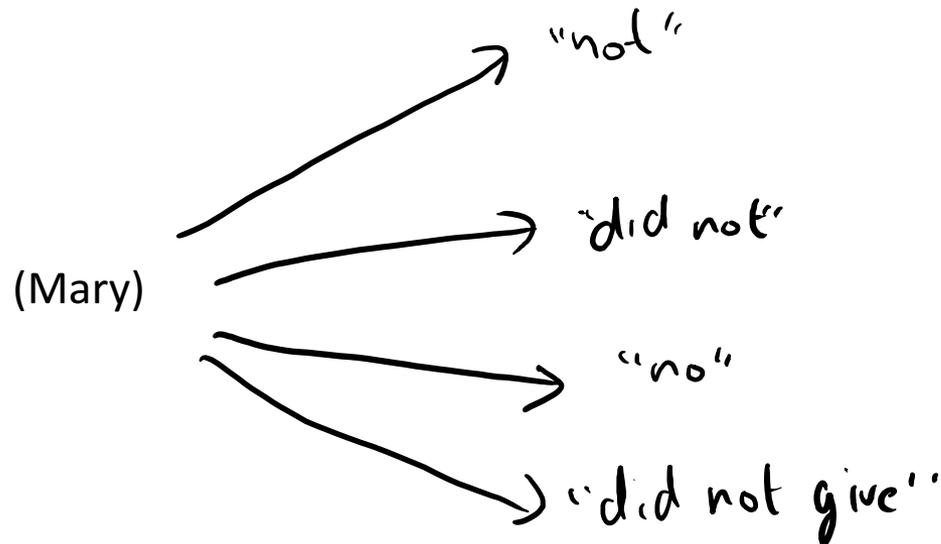
h
 $E: ""$
 $F: - - - -$
 $C: 0.0$

h'

$E: "Mary"$
 $F: \alpha - - - -$
 $C: \log P(\text{Mary}|\text{Maria})^B = 1$
 $+ \log(\text{Mary}|\langle s \rangle, \langle s \rangle)$

Monotonic Phrase Decoding

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



Monotonic Phrase Decoding

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

(Mary) (did not)

Monotonic Phrase Decoding

Maria	no	dio	una	bofetada	a	la	bruja	verde
-------	----	-----	-----	----------	---	----	-------	-------

<u>Mary</u>	<u>not</u>	<u>give</u>	<u>a</u>	<u>slap</u>	<u>to</u>	<u>the</u>	<u>witch</u>	<u>green</u>
	<u>did not</u>		<u>a</u>	<u>slap</u>	<u>by</u>		<u>green</u>	<u>witch</u>
	<u>no</u>		<u>slap</u>		<u>to the</u>			
	<u>did not give</u>				<u>to</u>			
				<u>slap</u>		<u>the</u>		
							<u>the</u>	<u>witch</u>

(Mary) (did not) (slap)

Monotonic Phrase Decoding

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

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

Monotonic Phrase Decoding

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

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

Non-Monotonic Decoding

h E: $P_1 + P_2$ h'
F: * * - ↓ *
C:

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

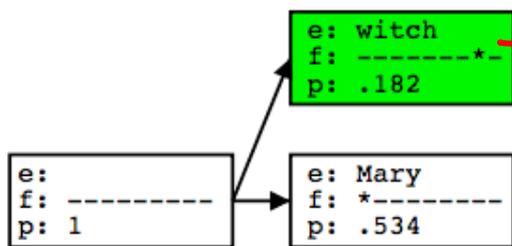
Maria	no	dio	una	bofetada	a	la	bruja	verde
-------	----	-----	-----	----------	---	----	-------	-------

<u>Mary</u>	<u>not</u>	<u>give</u>	<u>a</u>	<u>slap</u>	<u>to</u>	<u>the</u>	<u>witch</u>	<u>green</u>
	<u>did not</u>		<u>a</u>	<u>slap</u>	<u>by</u>		<u>green</u>	<u>witch</u>
	<u>no</u>		<u>slap</u>		<u>to the</u>			
	<u>did not give</u>				<u>to</u>			
					<u>the</u>			
			<u>slap</u>			<u>the</u>	<u>witch</u>	

```
e:
f: -----
p: 1
```


Non-Monotonic Decoding

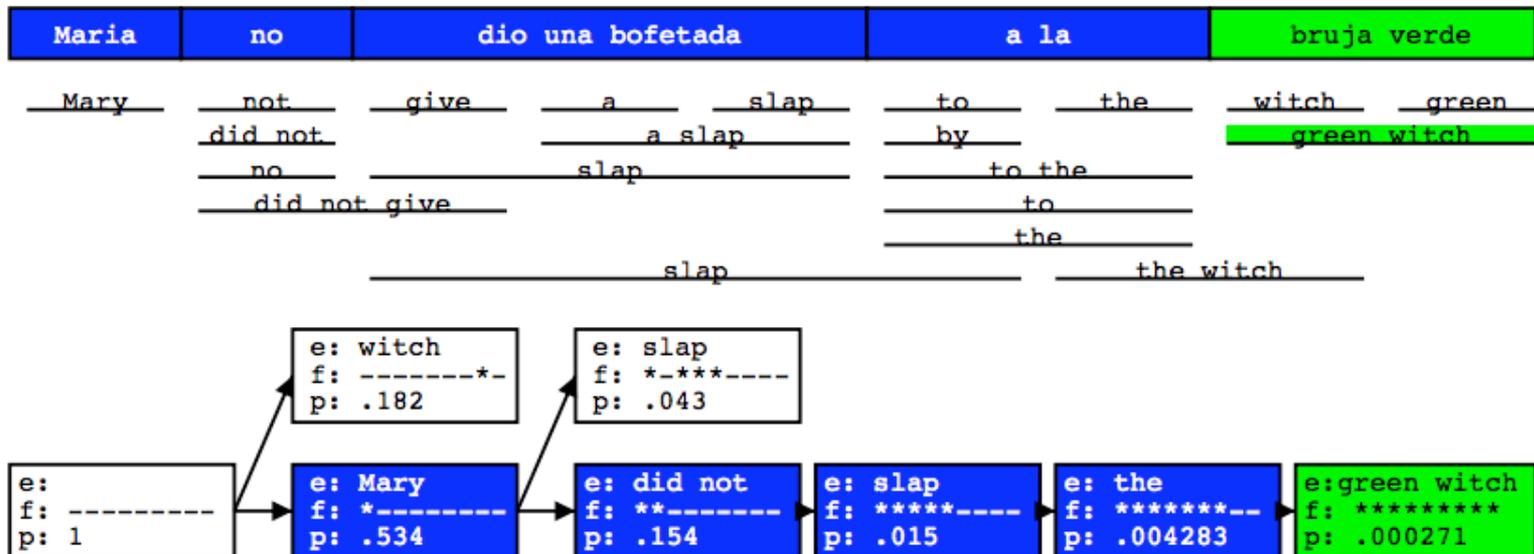
Maria	no	dio	una	bofetada	a	la	bruja	verde
<u>Mary</u>	<u>not</u>	<u>give</u>	<u>a</u>	<u>slap</u>	<u>to</u>	<u>the</u>	<u>witch</u>	<u>green</u>
	<u>did not</u>		<u>a</u>	<u>slap</u>	<u>by</u>		<u>green</u>	<u>witch</u>
	<u>no</u>		<u>slap</u>		<u>to the</u>			
	<u>did not give</u>				<u>to</u>			
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



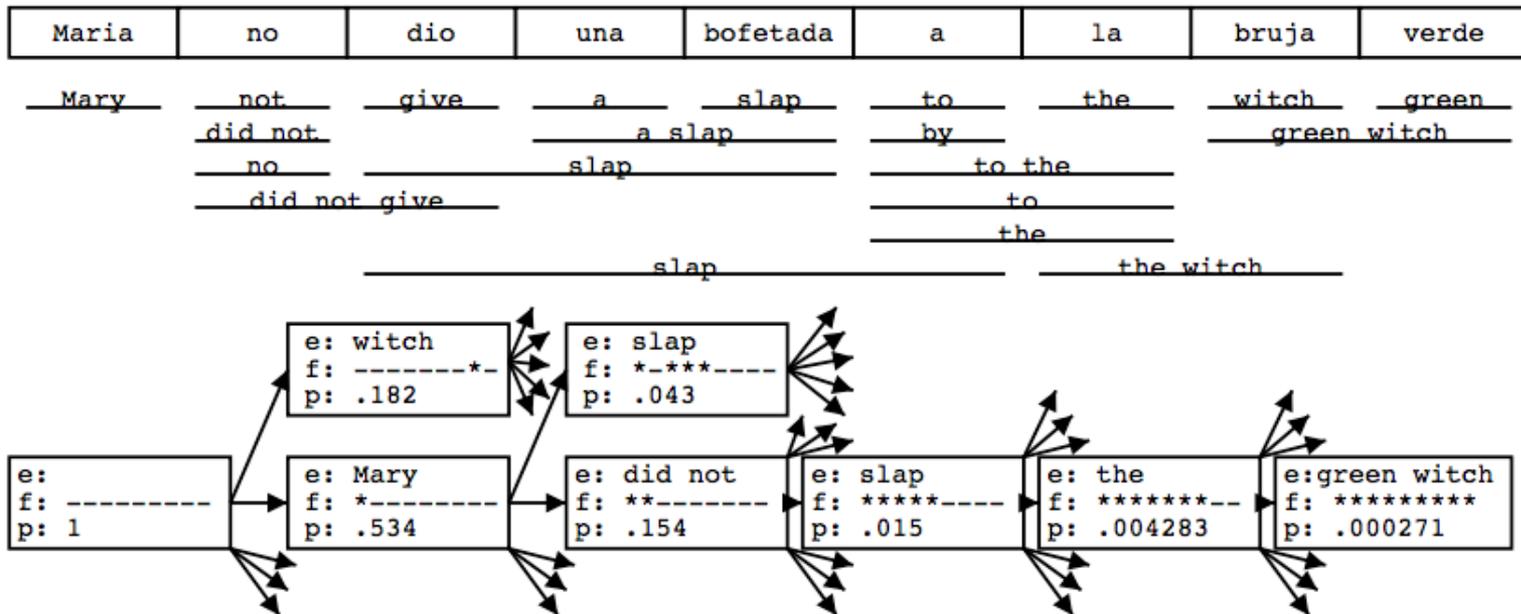
e: witch Mary
*f: *-----**

e: witch slap
f: _-++*+*+_-*

Non-Monotonic Decoding



Non-Monotonic Decoding

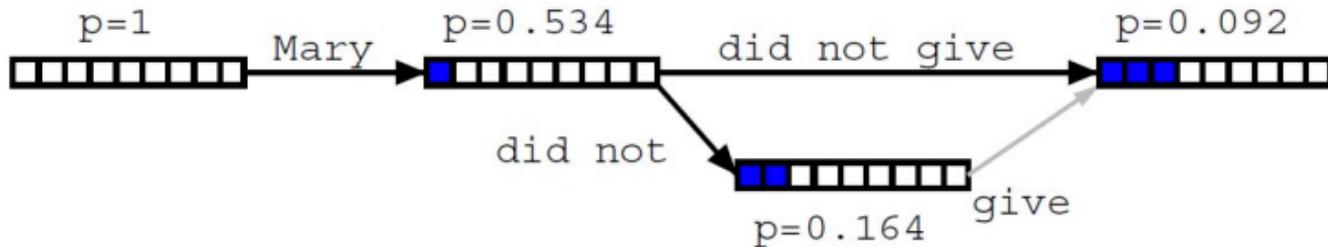
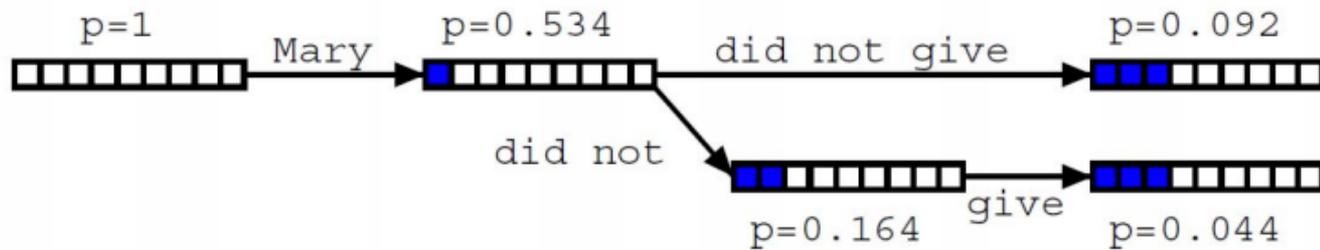


Comparing Partial Translations

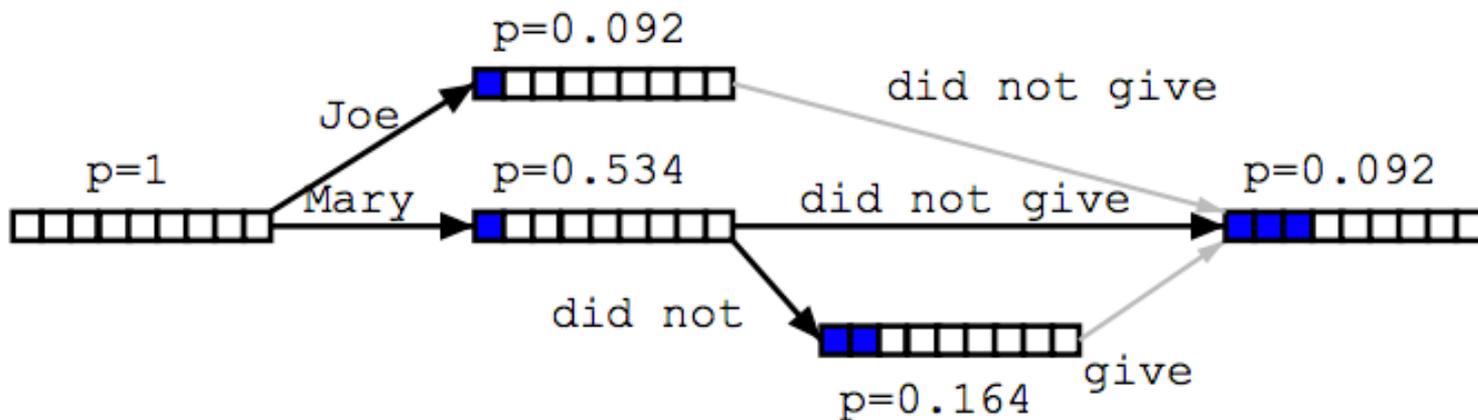
h_1 E: "Mary"
F: * _ _ _
C:

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 $i = 0 \dots n-1$

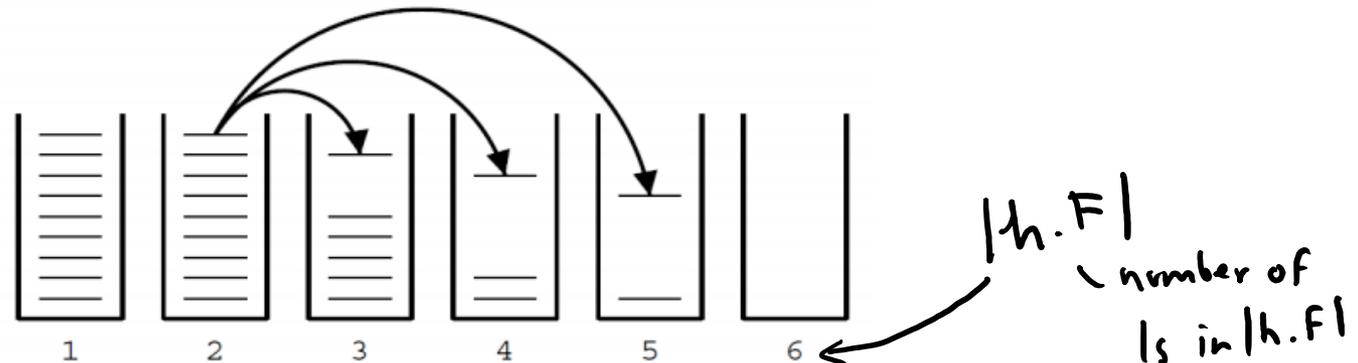
for h in k-best hypothesis from stack[i]

for h' in possible expansion of h

assign a score to h'

push h' onto stack[$\text{len}(h')$]

return $\text{max}(\text{stack}[n])$



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

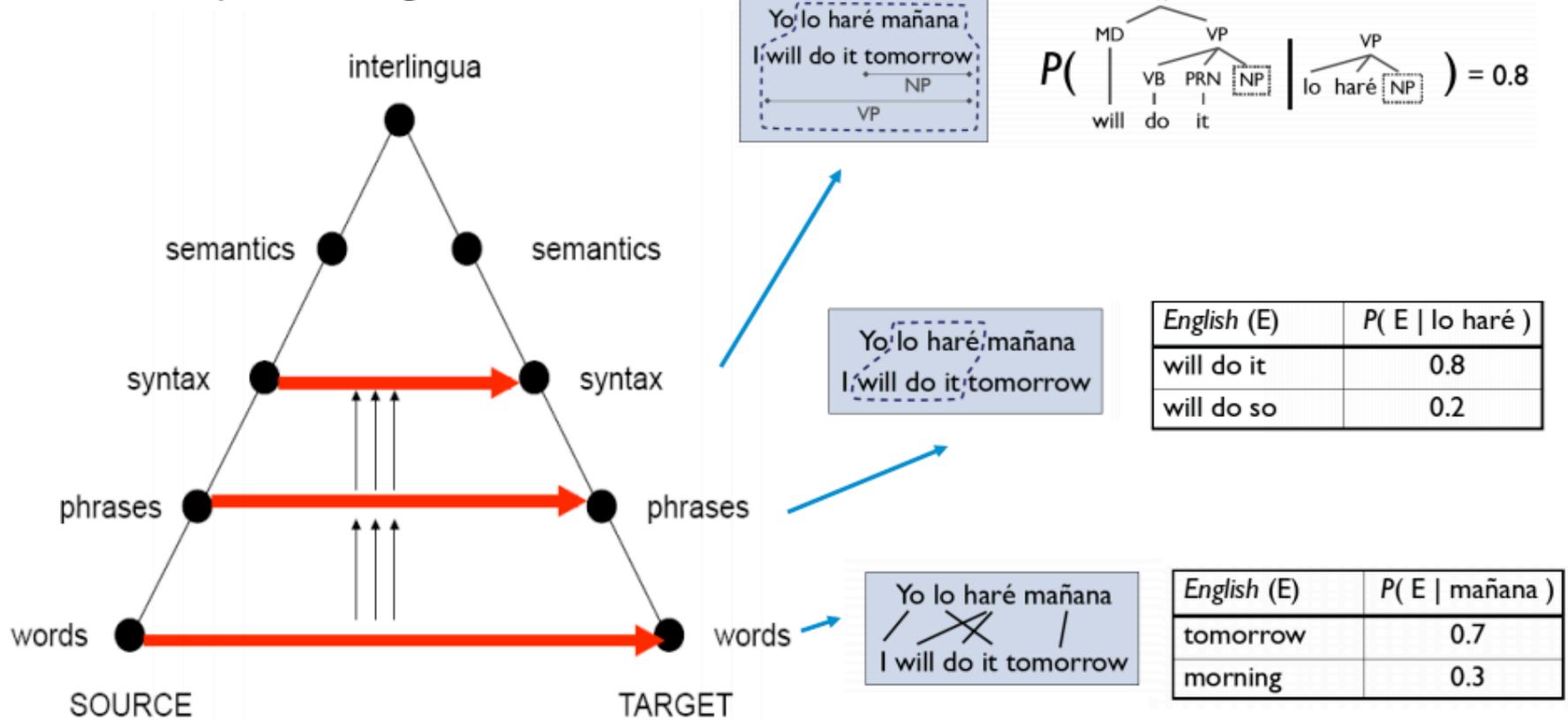
The draft approves the Parliament

Syntax leads to Good English

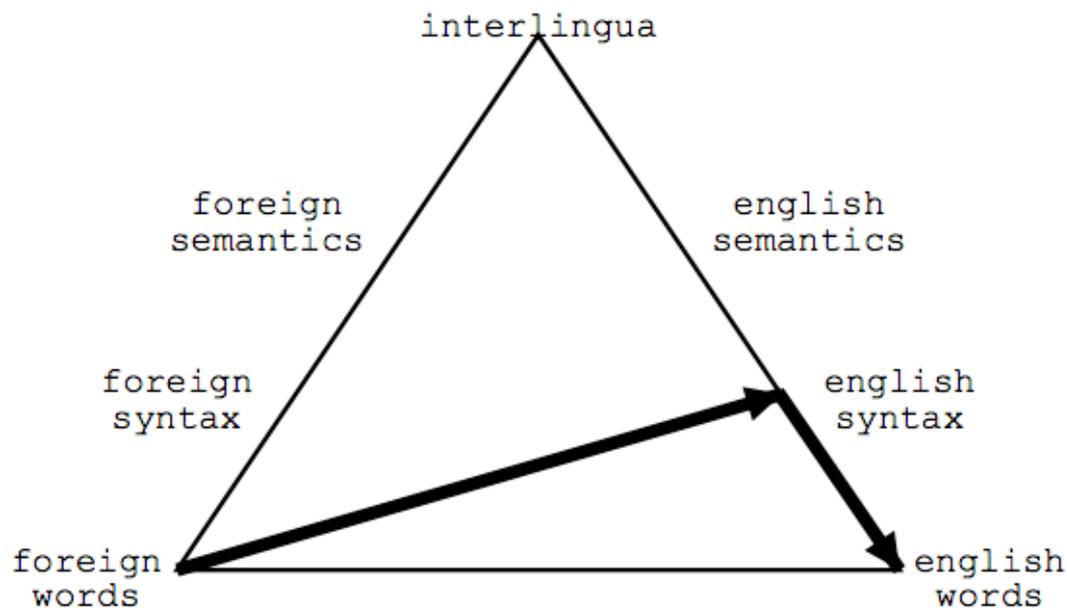
the house of the man is small the house **is** the man is small

The Vauquios Triangle

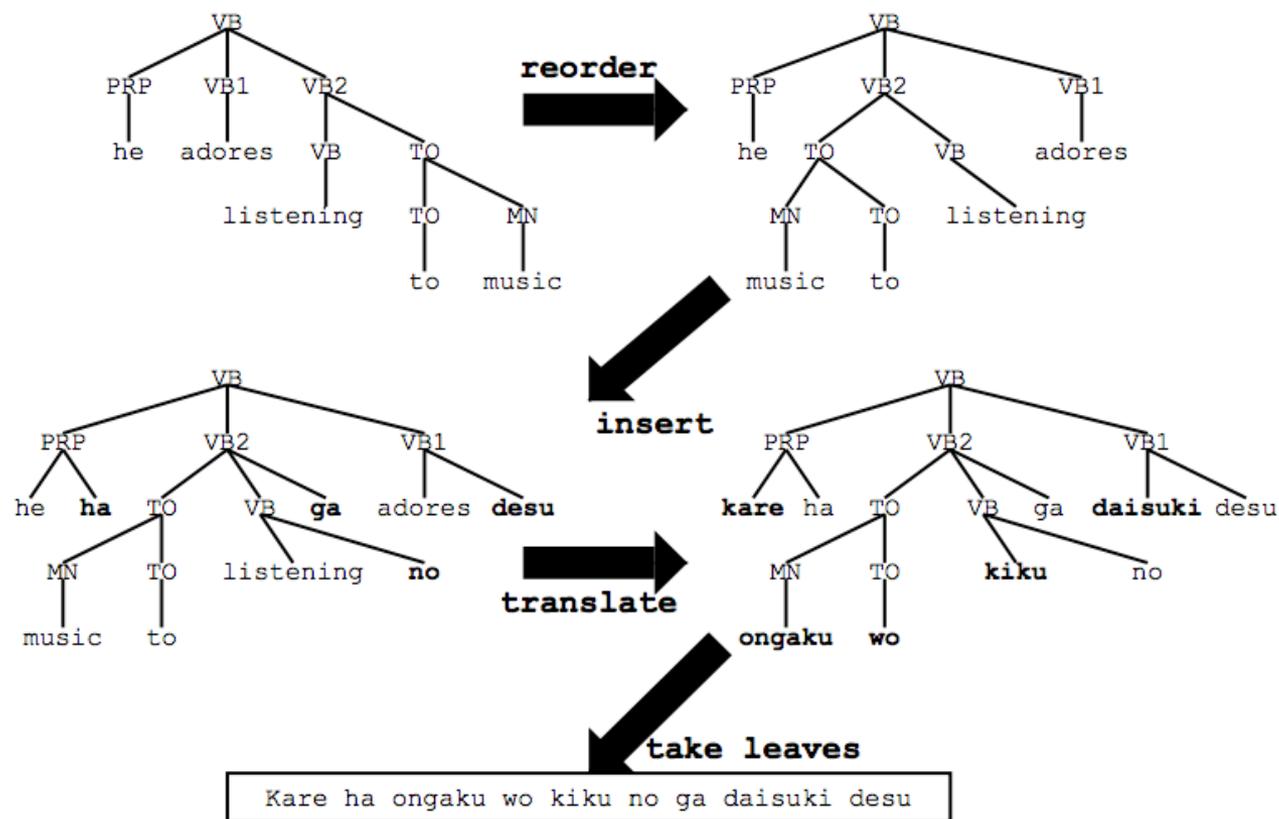
“Vauquios Triangle”



String to Tree Translation

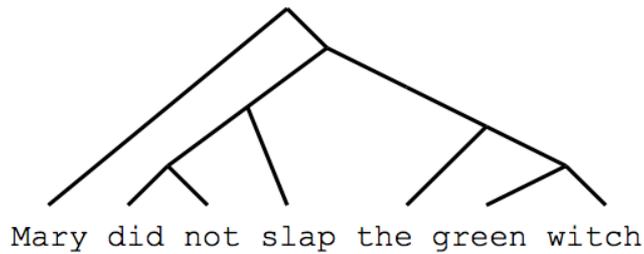


String to Tree Translation



[from Yamada and Knight, 2001]

Synchronous CFGs



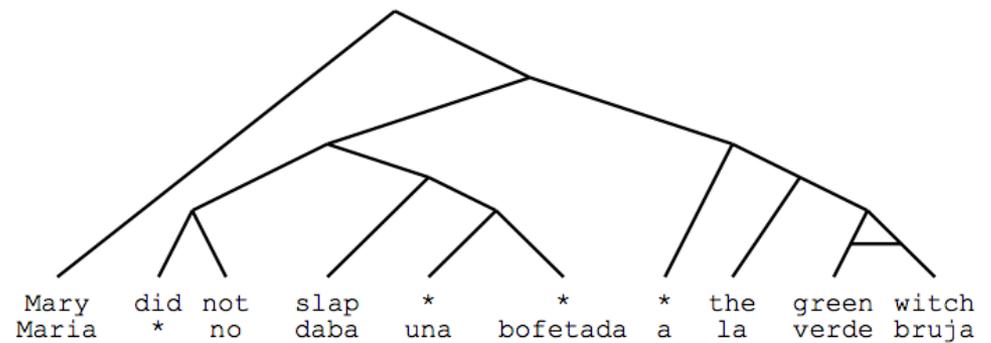
$$A \rightarrow A_1 A_2 \parallel A_1 A_2$$

$$A \rightarrow A_1 A_2 \parallel A_2 A_1$$

$$A \rightarrow e \parallel f$$

$$A \rightarrow e \parallel *$$

$$A \rightarrow * \parallel f$$



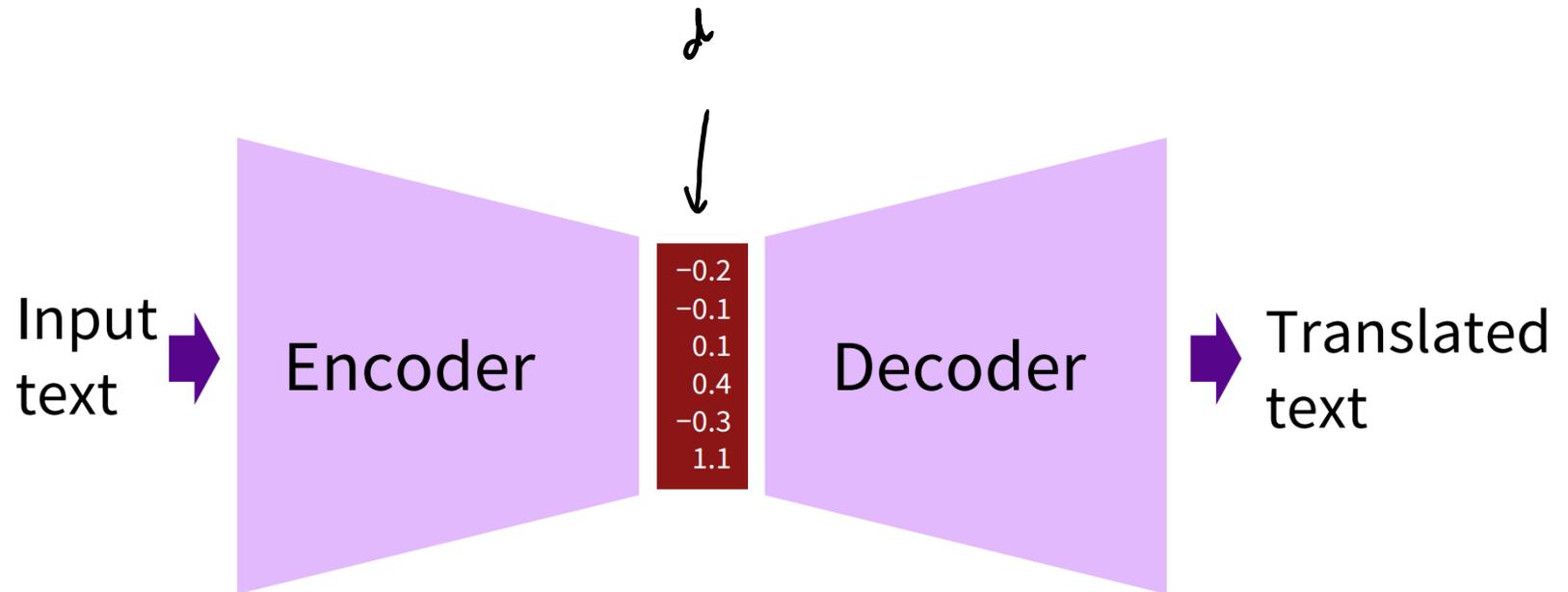
Outline

Decoding Algorithms

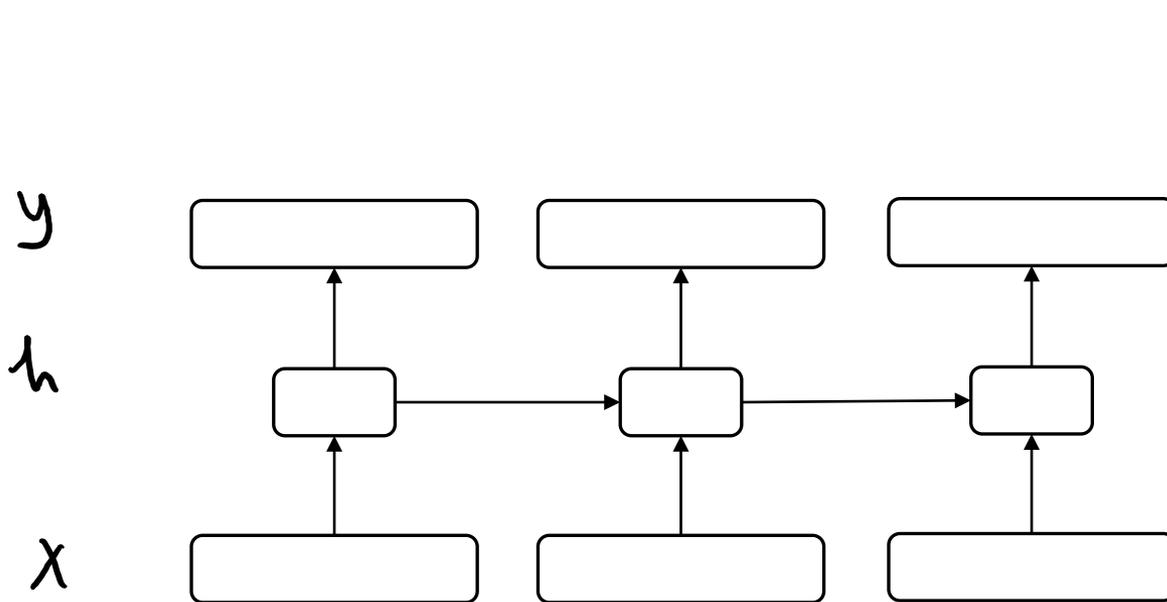
Syntax-Based MT

Neural MT Models

Neural MT Models

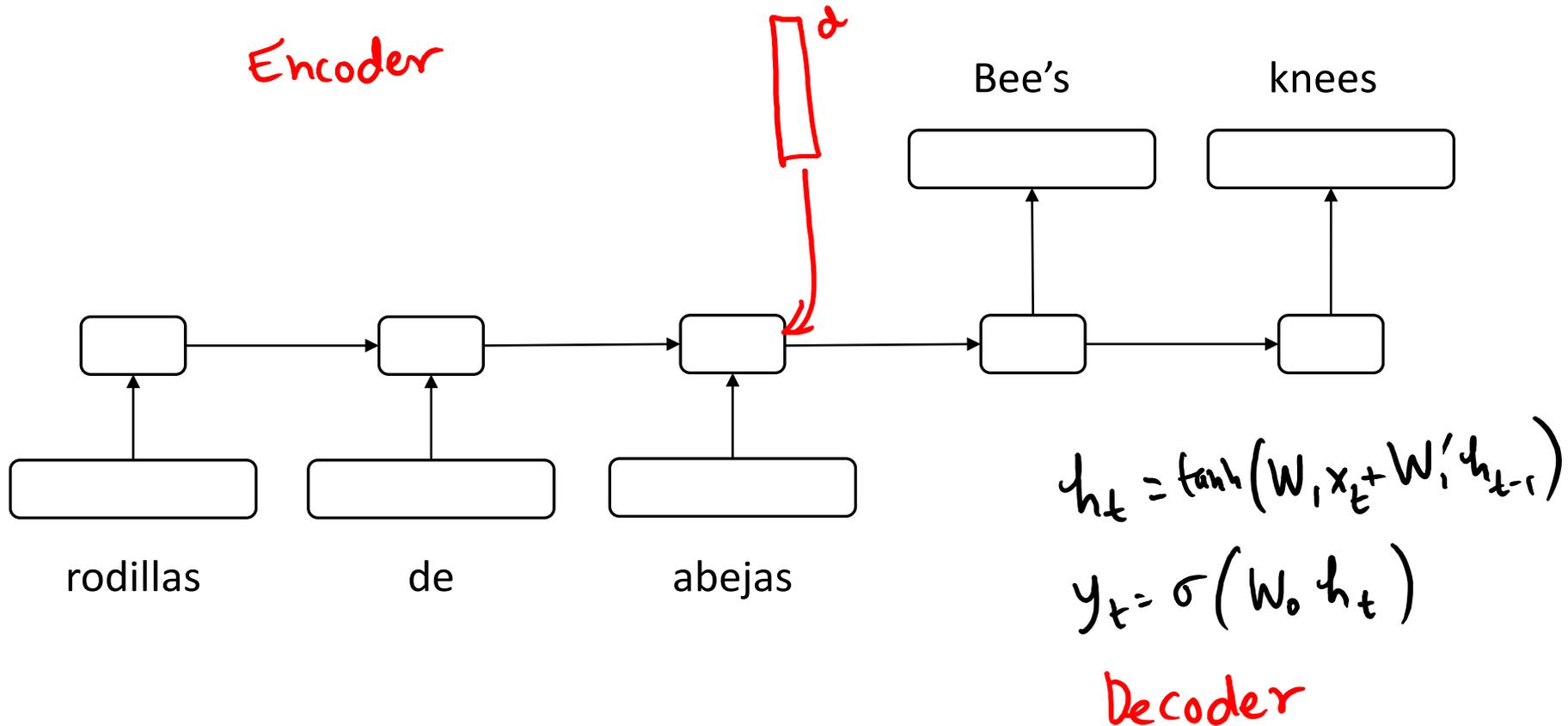


Recurrent Neural Networks



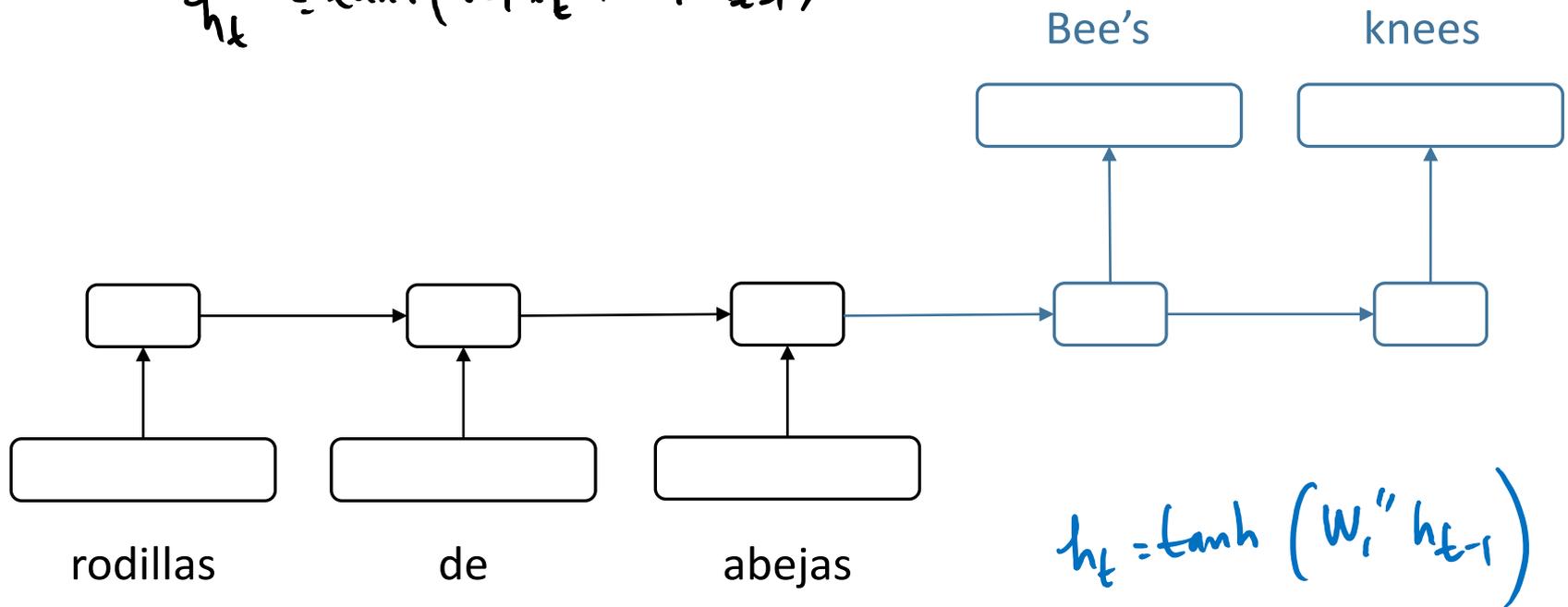
$$y_t = \text{softmax}(W_o h_t)$$
$$h_t = \text{tanh}(W_i x_t + W_i' h_{t-1})$$

Recurrent Neural Networks

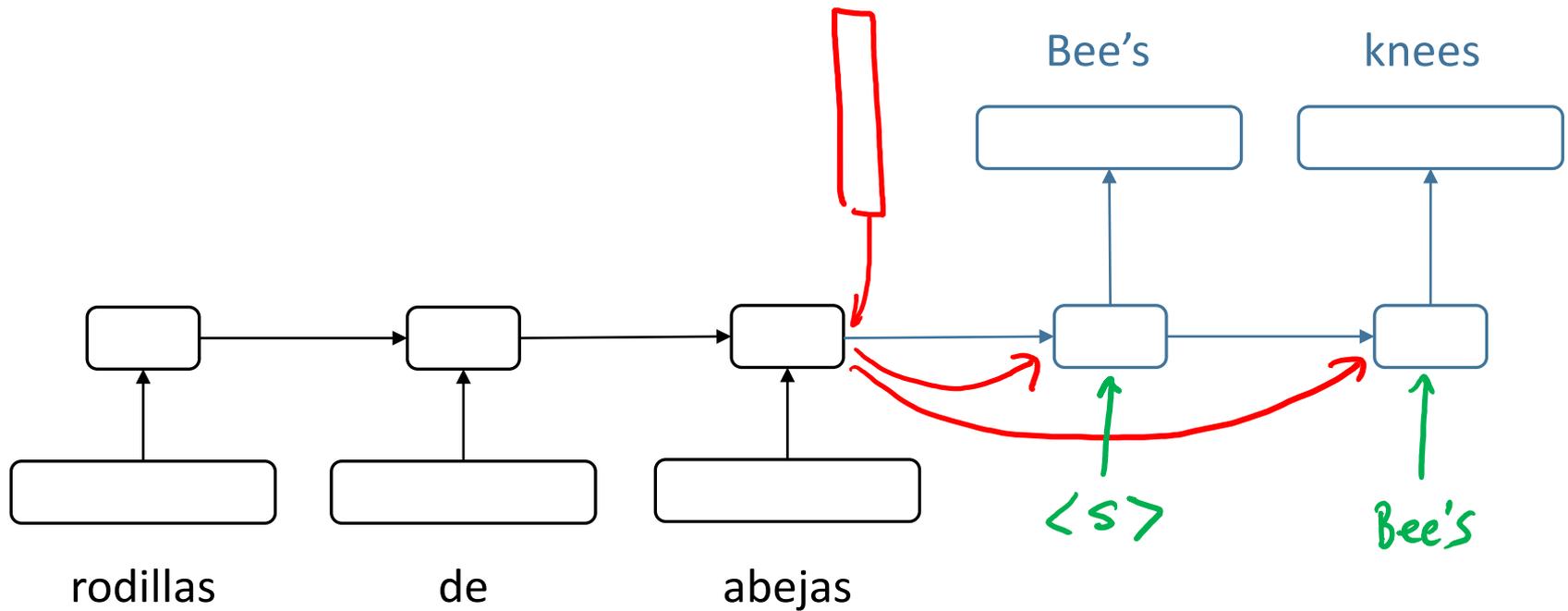


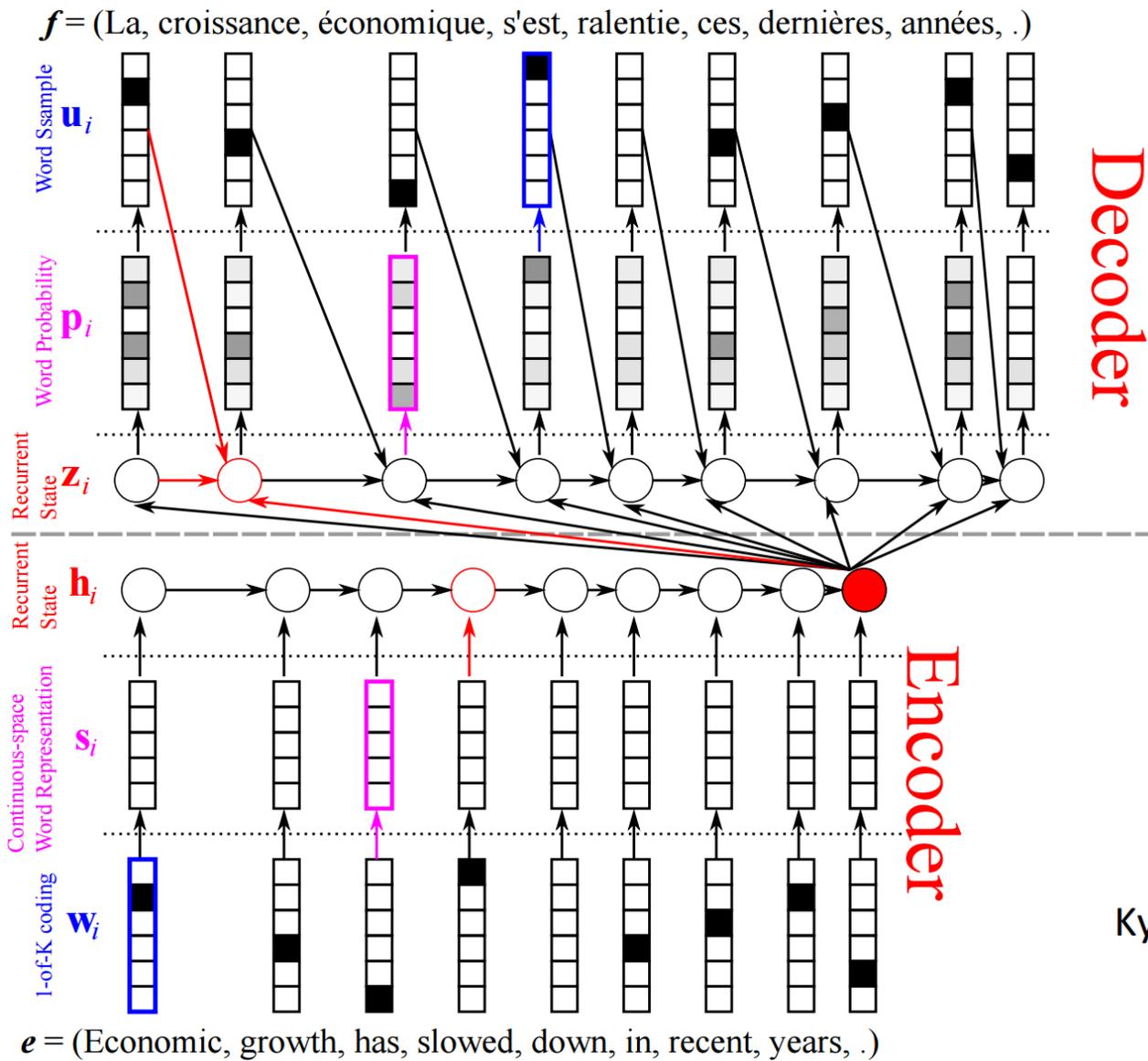
Different Weights

$$h_k = \tanh(W_k x_k + W_k' h_{k-1})$$



More Connections

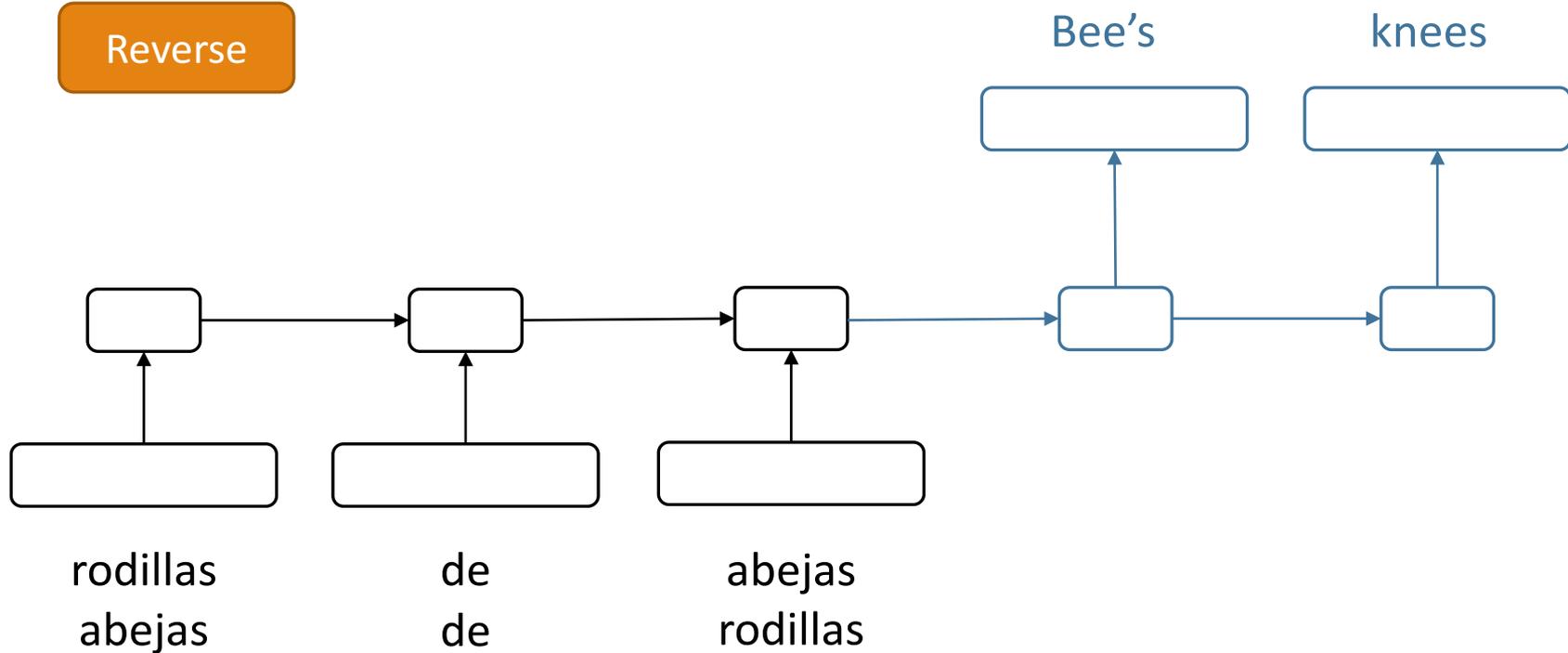




Kyunghyun Cho et al. 2014

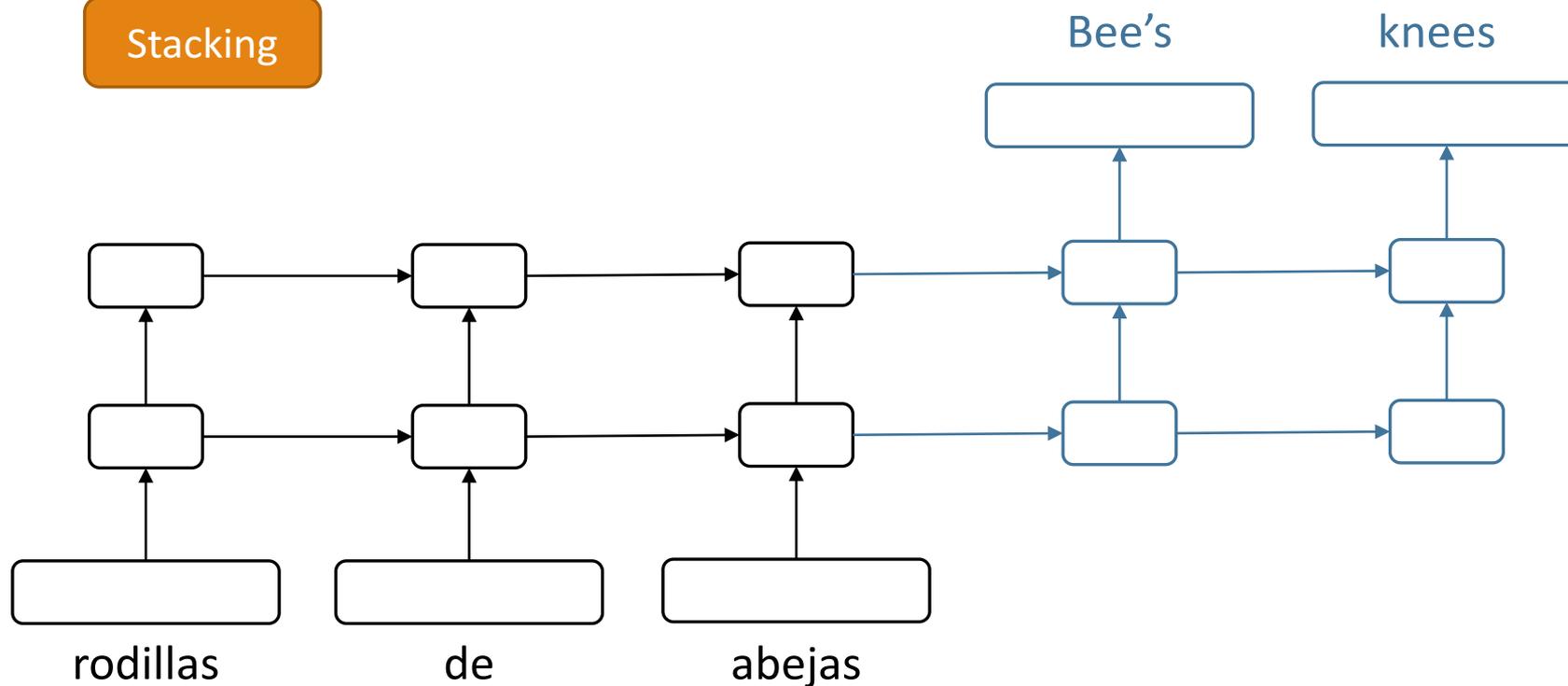
Other Extensions

Reverse



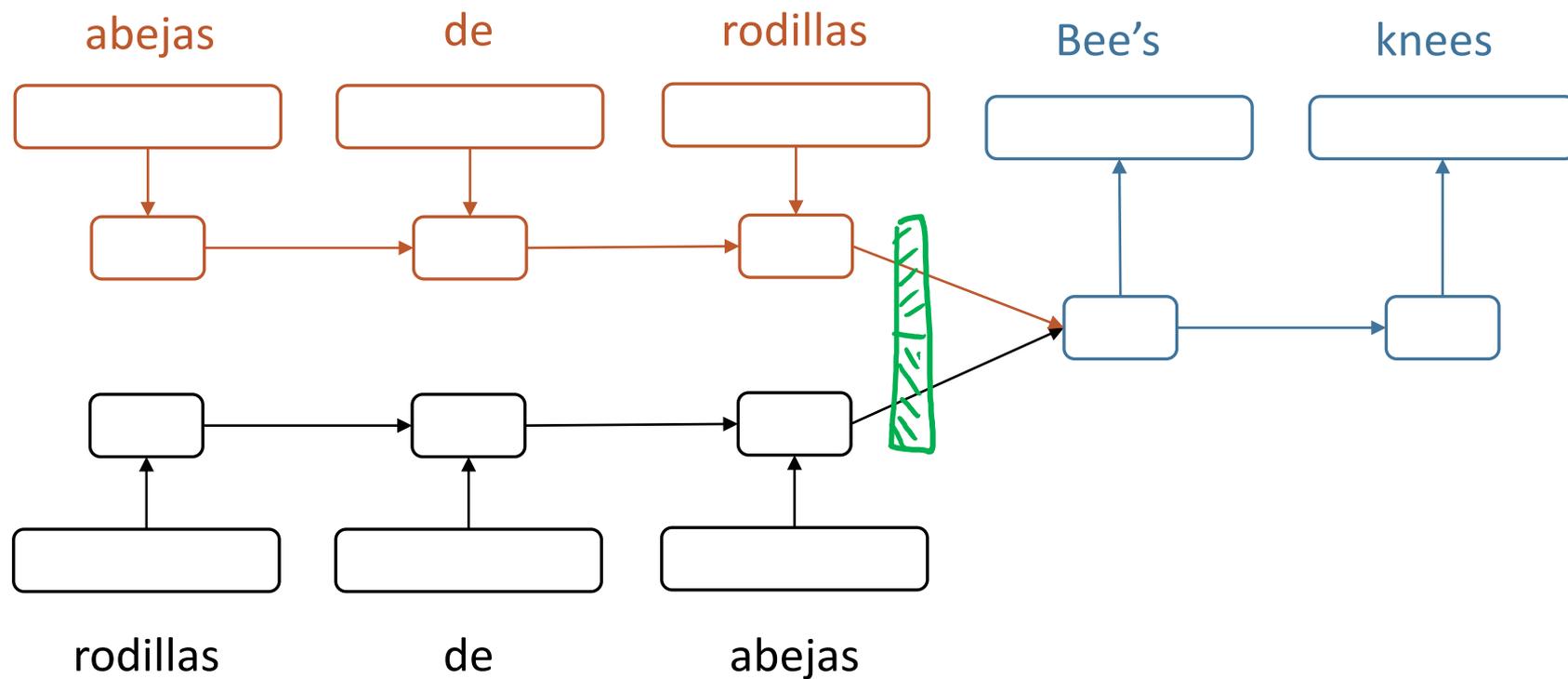
Other Extensions

Stacking

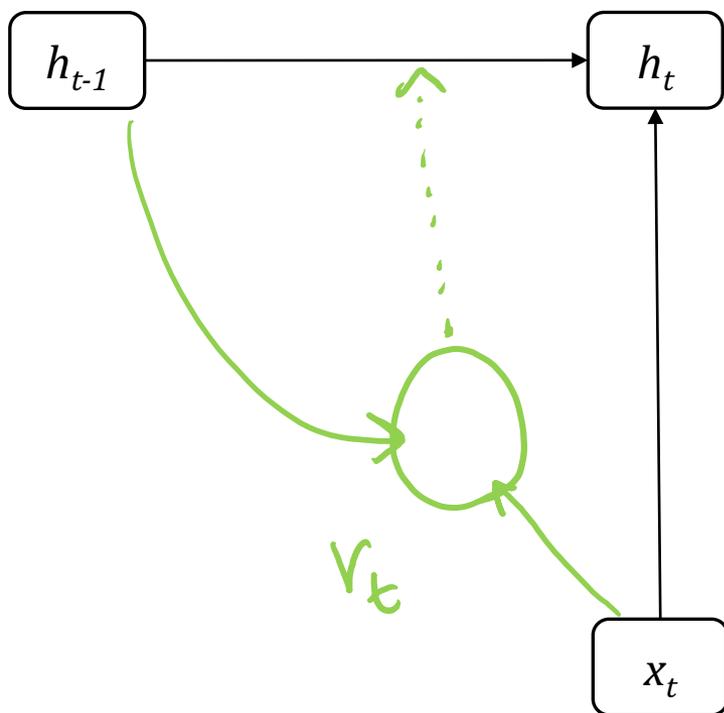


Other Extensions

Bi-directional



Regular Recurrent Units

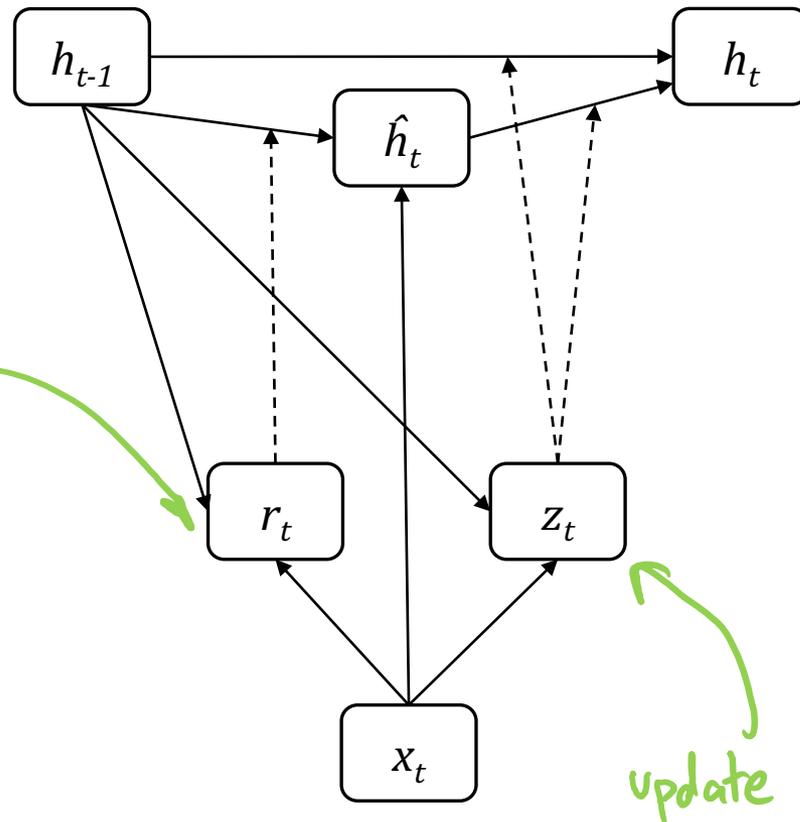


$$h_t = \tanh(W_1 x_t + W_1' h_{t-1})$$

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

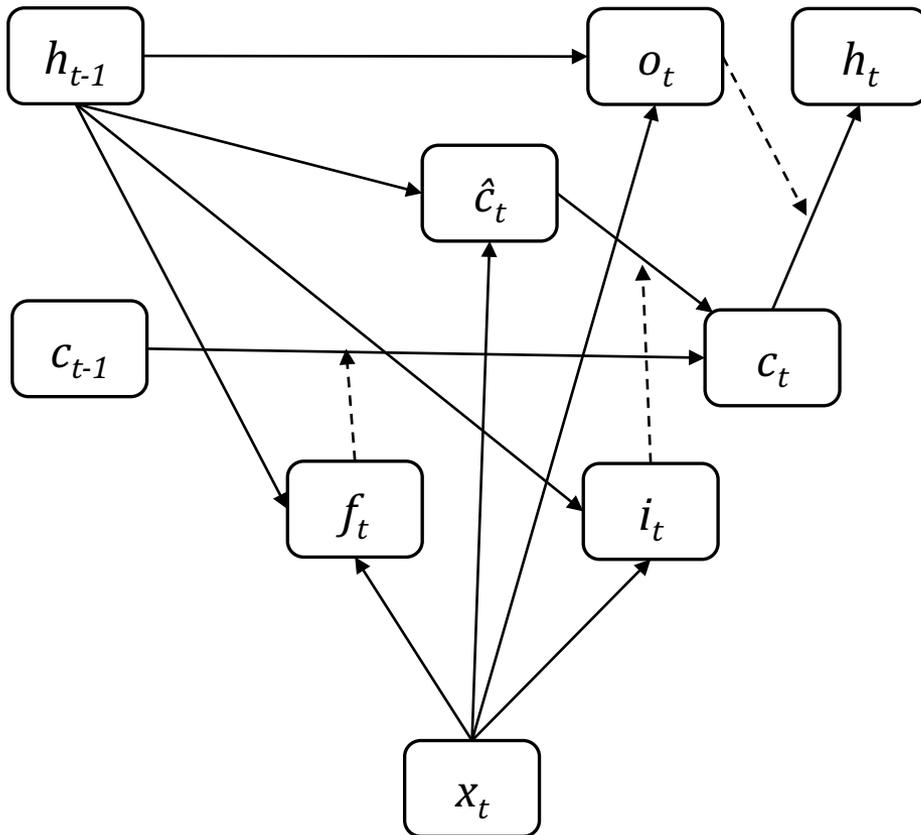
$$v_t = \text{sigmoid}(W_2 x_t + W_2' h_{t-1})$$

Gated Recurrent Units



$$z_t = \sigma \left(W^{(z)} x_t + U^{(z)} h_{t-1} \right)$$
$$r_t = \sigma \left(W^{(r)} x_t + U^{(r)} h_{t-1} \right)$$
$$\tilde{h}_t = \tanh \left(W x_t + r_t \circ U h_{t-1} \right)$$
$$h_t = z_t \circ h_{t-1} + (1 - z_t) \circ \tilde{h}_t$$

Long Short-Term Memory



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

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

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

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

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

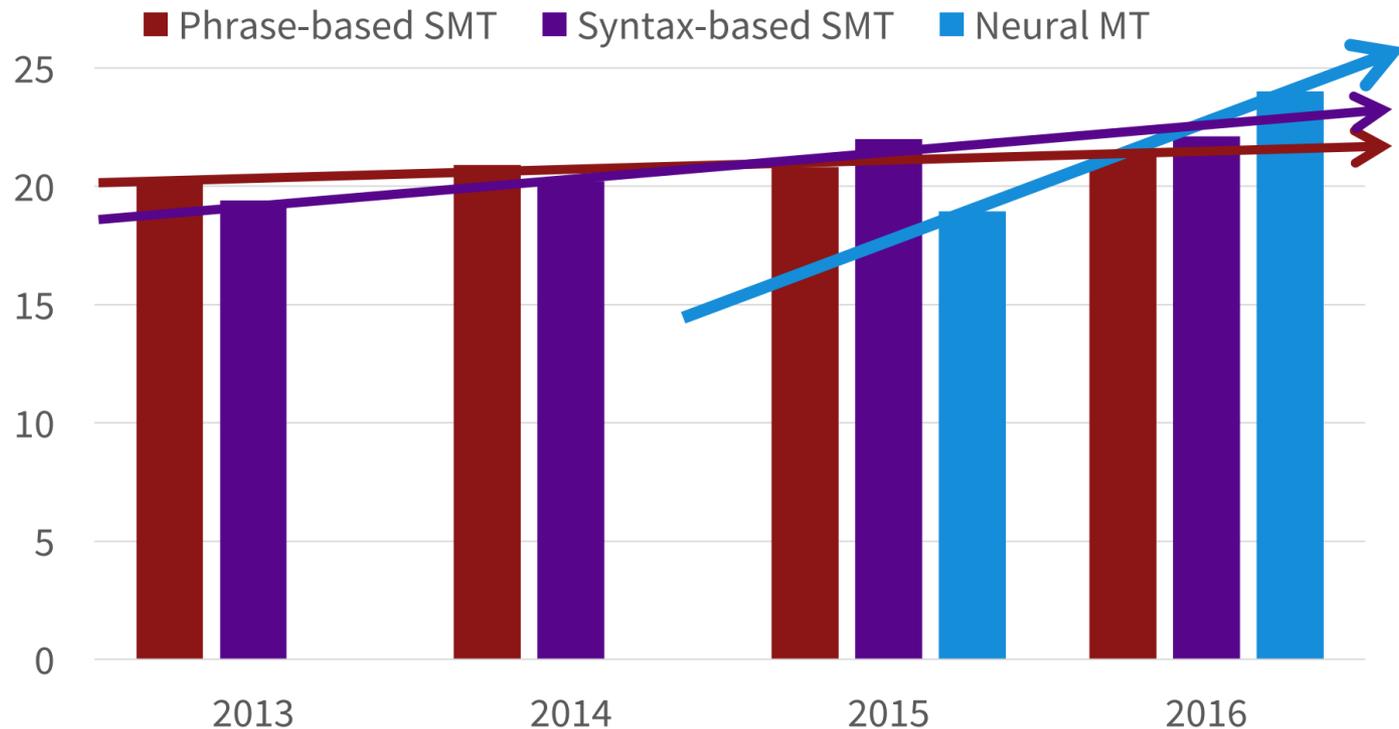
$$h_t = o_t \circ \tanh(c_t)$$

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

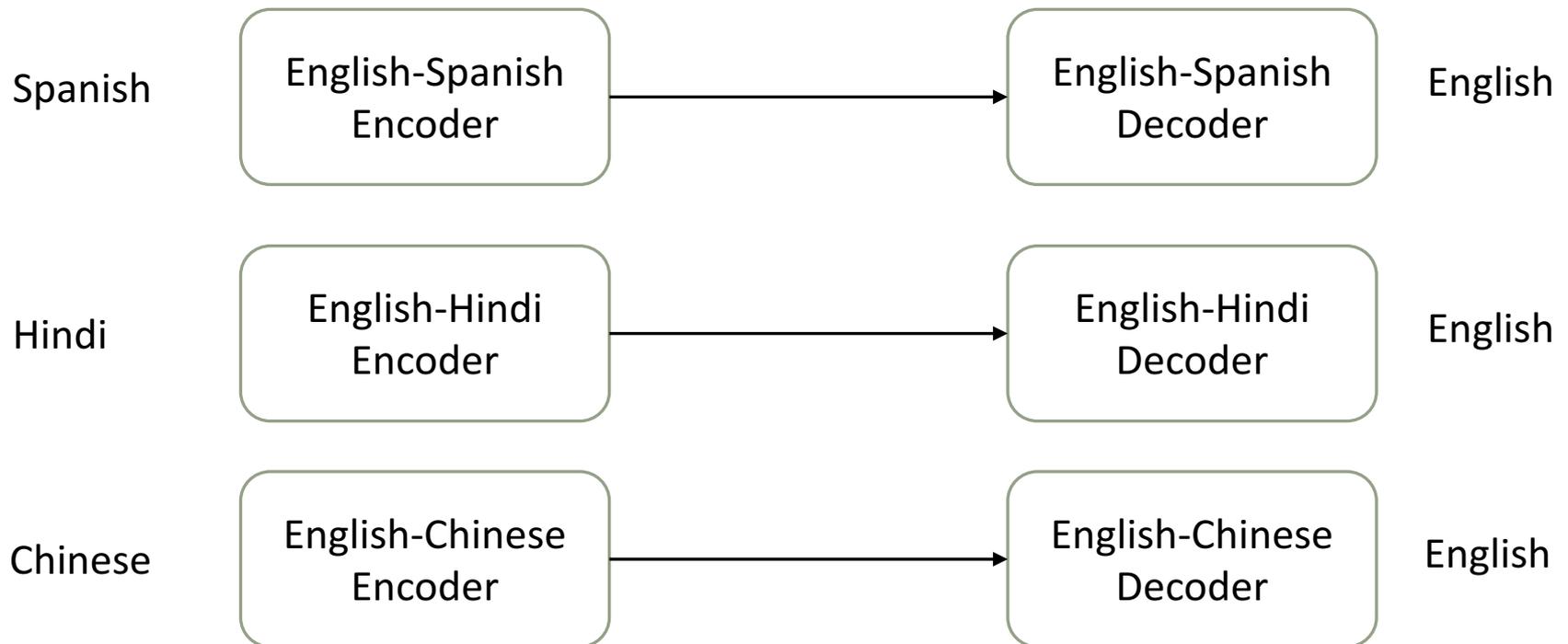
Sequence to Sequence Learning by Sutskever et al. 2014

Trend in Machine Translation

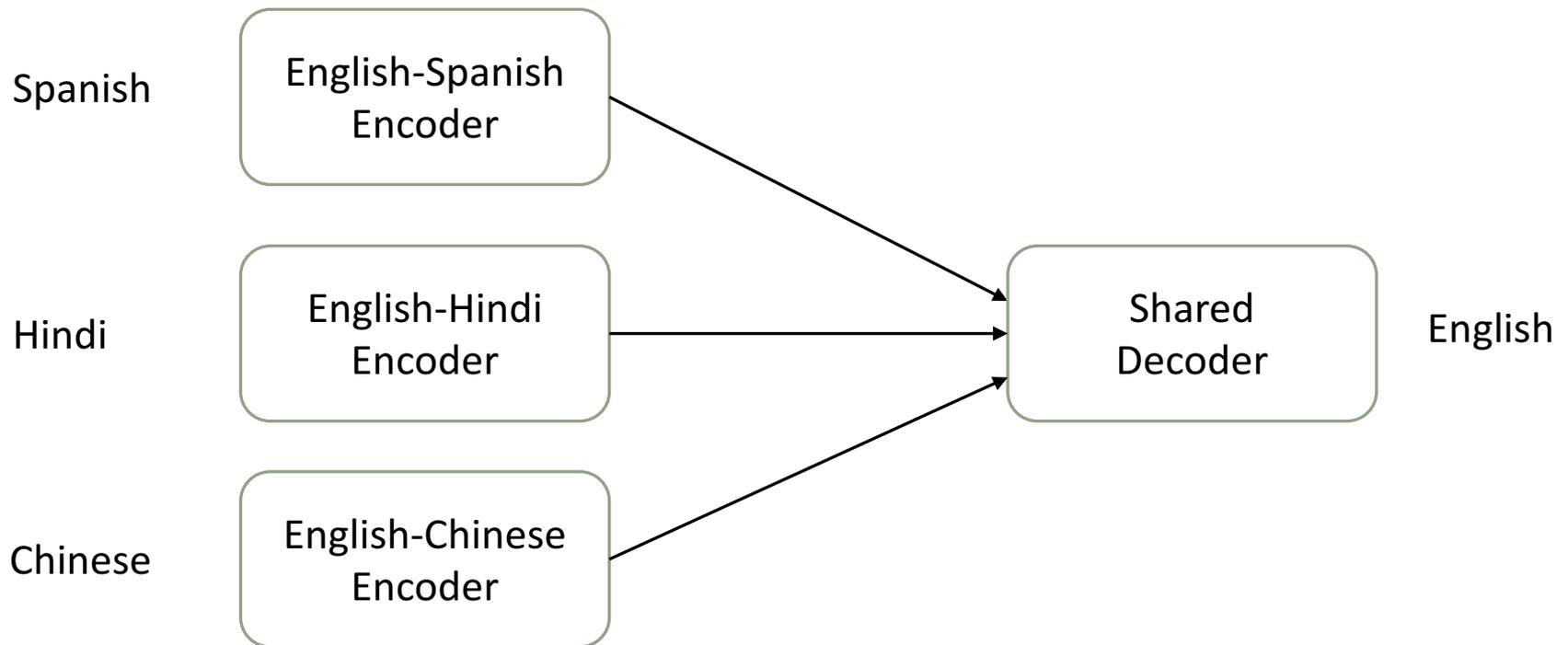


From [Sennrich 2016, http://www.meta-net.eu/events/meta-forum-2016/slides/09_sennrich.pdf]

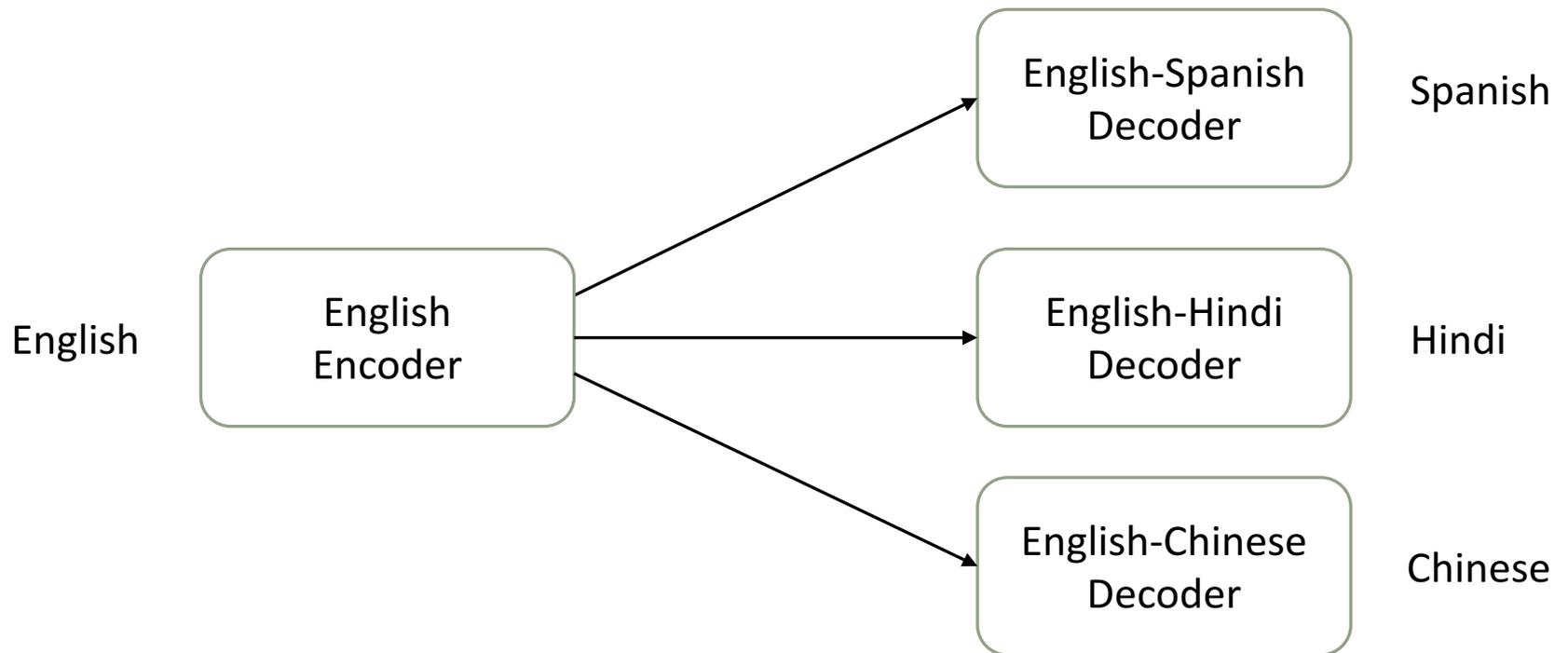
Multilingual Neural MT: Naïve



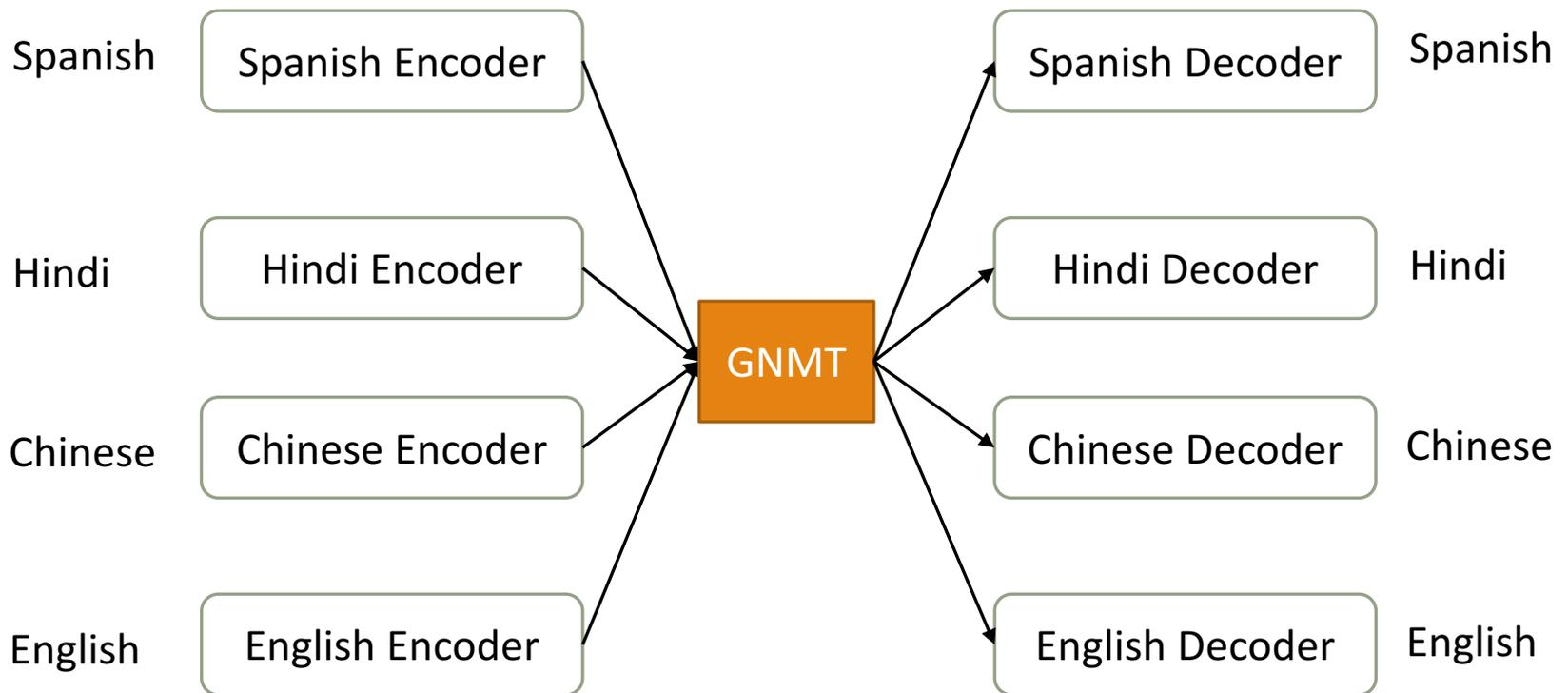
Multilingual NMT: Decoder



Multilingual NMT: Encoder



Google Neural MT



Out of Control!



Out of Control!

Google Sign in

Translate Turn off instant translation

Russian German English Detect language ↔ English German Spanish Translate

Deutschland

Deutsch, deutsch, deutsch, deutsch, deutsch, deutsch

Natürlich hat ein Deutscher "Wetten, dass ... ?" erfunden
Vielen Dank für die schönen Stunden!
Wir sind die freundlichsten Kunden auf dieser Welt
Wir sind bescheiden, wir haben Geld
Die Allerbesten in jedem Sport
Die Steuern hier sind Weltrekord
Bereisen Sie Deutschland und bleiben Sie hier!
Auf diese Art von Besuchern warten wir
Es kann jeder hier wohnen, dem es gefällt
Wir sind das freundlichste Volk auf dieser Welt

Deutsch, deutsch, deutsch, deutsch

Germany

German, English, German, German, German, and English

Of course a German has "betting that ...?" invented
Thanks for the nice hours!
We are the friendliest customers in this world
We are modest, we have money
The very best in any sport
The taxes here are a world record
Travel to Germany and stay here!
We are waiting for this kind of visitors
Anyone who likes it can live here
We are the friendliest people in this world

English, German, German, and German

Out of Control!

The image displays two examples of a bug in Google Translate. In the first example, the source text is "knife, fork, knife,". The target text is "Messer, Messer, Messer,". A red annotation reads "(The trailing comma messes this one up.)". In the second example, the source text is "Messer, Gabel, Messer, Messer, Messer, Messer, Messer, Messer, Messer, Messer, Messer". The target text is "Screen monitor styling Projector styling Print styling ← back to. 2010-01-20 with adjustable interlinear. Knife, fork; knife, knife, knife, knife;".

English German Spanish Detect language ▾ ↕ German English Spanish ▾ Translate

knife, fork, knife, ✕
(The trailing comma messes this one up.)
19/5000

German English Spanish ▾ Translate

Messer, Gabel, Messer, Messer, Messer, Messer, Messer, Messer, Messer, Messer, Messer ✕
77/5000

German English Spanish ▾ Translate

Screen monitor styling Projector styling Print styling ← back to. 2010-01-20 with adjustable interlinear. Knife, fork; knife, knife, knife, knife;