

Machine Translation Contd

Prof. Sameer Singh

CS 295: STATISTICAL NLP

WINTER 2017

March 2, 2017

Omer Levy



AI/ML Seminar
Monday, March 6th
1pm-2pm
DBH 4011

Understanding Word Embeddings

Meeting with Graduate Students
4:00-4:45pm
Room TBA (email me)

Upcoming...

Project

- Status report due in 1 weeks: **March 7, 2017**
- Instructions coming today!
- Almost final report, only **5 pages**

Summaries

- Paper summaries: **March 14**
- Summary 1 graded

Homework

- Homework 4 is due on **March 13**
- Write-up and data releasing soon.

Outline

EM-Algorithm for Alignments

Phrase-Based MT

Decoding Algorithms

Syntax-Based MT

Outline

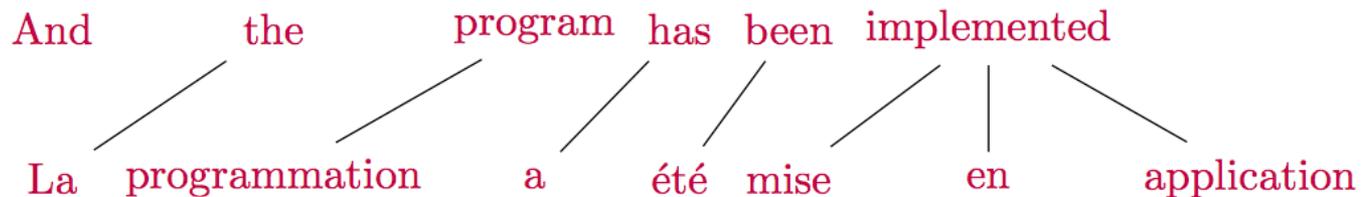
EM-Algorithm for Alignments

Phrase-Based MT

Decoding Algorithms

Syntax-Based MT

Parameters of the IBM Models



$$p(f|e) = \sum_a P(f, a|e)$$

$$\begin{aligned} P(f, a|e) = & q(2, 1, 6, 7) \times t(\text{La}|\text{the}) \\ & \times q(3|2, 6, 7) \times t(\text{Programmation}|\text{program}) \\ & \times q(4|3, 6, 7) \times t(\text{a}|\text{has}) \\ & \times q(5|4, 6, 7) \times t(\text{été}|\text{been}) \\ & \times q(6|5, 6, 7) \times t(\text{mise}|\text{implemented}) \\ & \times q(6|6, 6, 7) \times t(\text{en}|\text{implemented}) \\ & \times q(6|7, 6, 7) \times t(\text{application}|\text{implemented}) \end{aligned}$$

Parameters of the IBM Models

$$P(f, a | e) = \prod_i \underbrace{q(j | i, e, m)}_{\text{alignment}} \underbrace{t(f_i | e_j)}_{\text{translation}}$$

Translation from Alignments

qs

$$t(a|b) = \frac{c(a \rightarrow b)}{c(b)}$$

$$t(\text{"programme"}|\text{"program"}) = \frac{c(\text{"programme"} \rightarrow \text{"program"})}{c(\text{"program"})}$$

Alignments from Translation

$$P(a|e,f) = \frac{P(f|e,a) P(a)}{\sum_{a'} P(f|e,a') P(a')}$$

Expectation Maximization

Expectation

fixing θ \longrightarrow Getting z

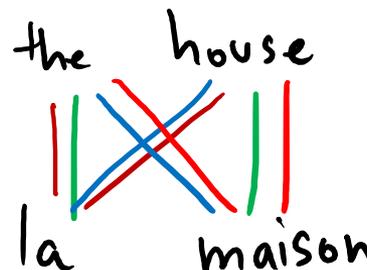
Maximization

fixing z \longrightarrow Getting θ

Example

► Translation probabilities

	the	house
la	0.4	0.1
maison	0.1	0.6



► Alignments

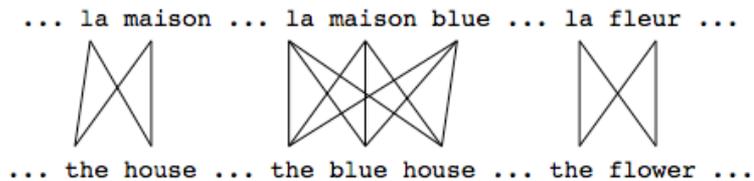
	$P(f, a e)$	$P(a f, e)$
the → la, house → la	$(0.4 \times 0.1) = 0.04$	~ 0.1
the → la, house → maison	$(0.4 \times 0.6) = 0.24$	~ 0.7
the → maison, house → la	$(0.1 \times 0.1) = 0.01$	~ 0.05
the → maison, house → maison	$(0.1 \times 0.6) = 0.06$	~ 0.15

► Counts

	the	house
la	$0.7 + 0.1 = 0.8$	0.15
maison	$0.05 + 0.15 = 0.2$	0.85

Example

Step 1



Word-based MT: Problems

Multi-word Alignments

$$P(\text{daba una botefada} \mid \text{slap}) \\ = P(\text{daba} \mid \text{slap}) P(\text{una} \mid \text{slap}) \\ P(\text{botefada} \mid \text{slap})$$

Non-compositionality

faire le ménage
make the house

Phrasal Translations

la comida me gusta mucho
I like the food a lot

Outline

EM-Algorithm for Alignments

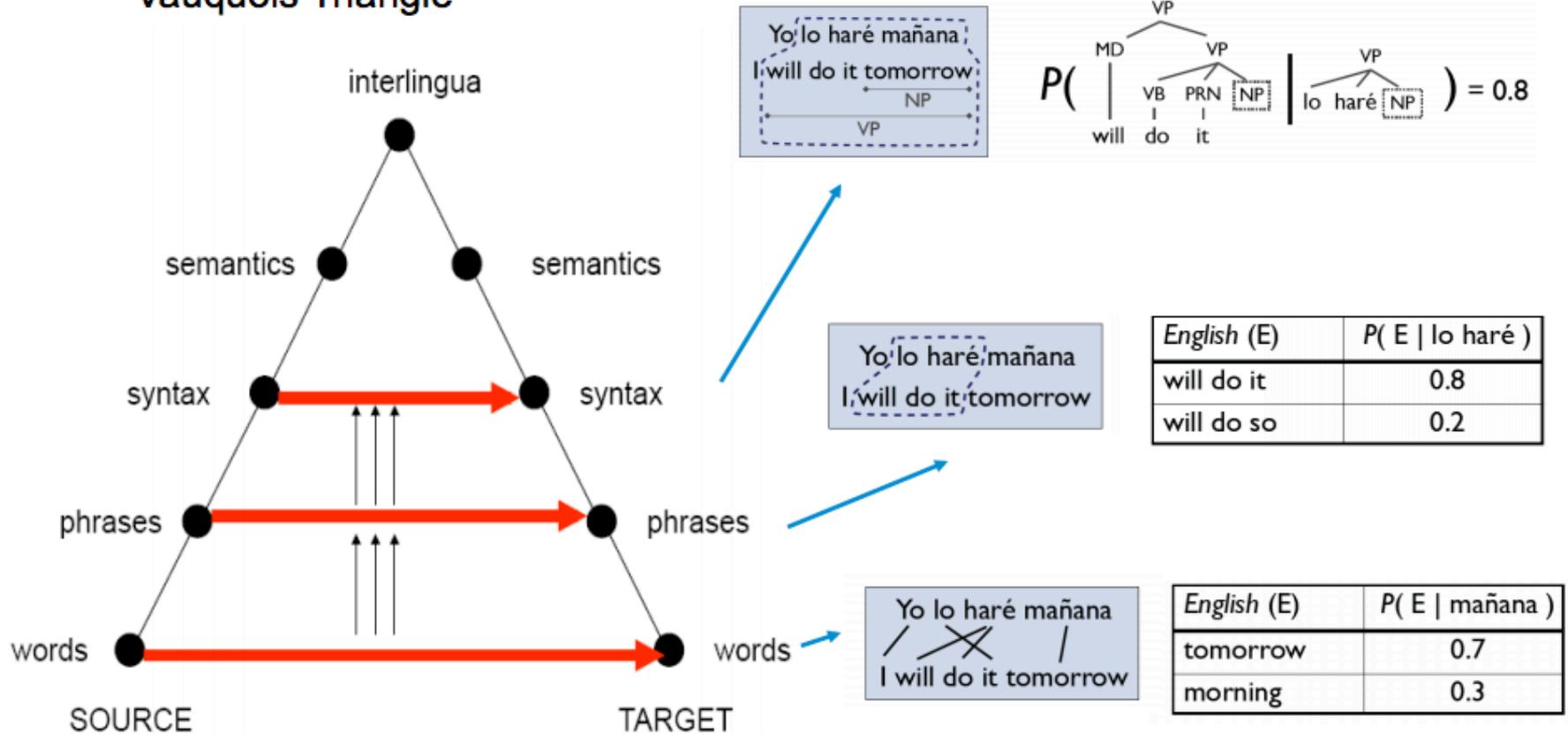
Phrase-Based MT

Decoding Algorithms

Syntax-Based MT

The Vauquios Triangle

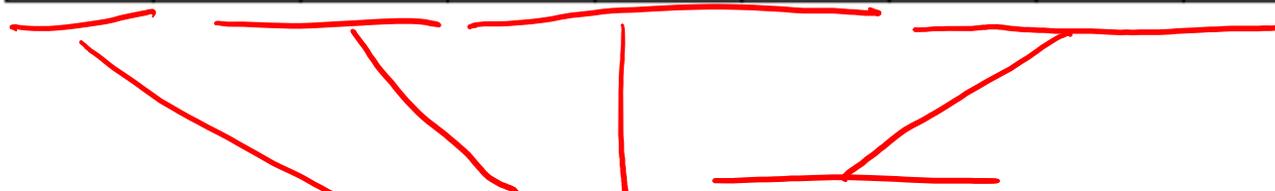
“Vauquios Triangle”



Phrase-based MT

$$P(f|e) P(e) \quad \prod_i P(f_i | e_i)$$
$$\prod_P P(f_P | e_P)$$

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



Mary did not slap the green witch

Phrase Lexicon

$$L = \{ (f, e, g) \}$$



$a u$ banque $to\ the\ bank$ 0.01

$allez\ au\ banque$ $go\ to\ the\ bank$ -2.5

Learning Phrasal Alignments

english to spanish

 bofetada bruja
Maria no daba una | a la | verde

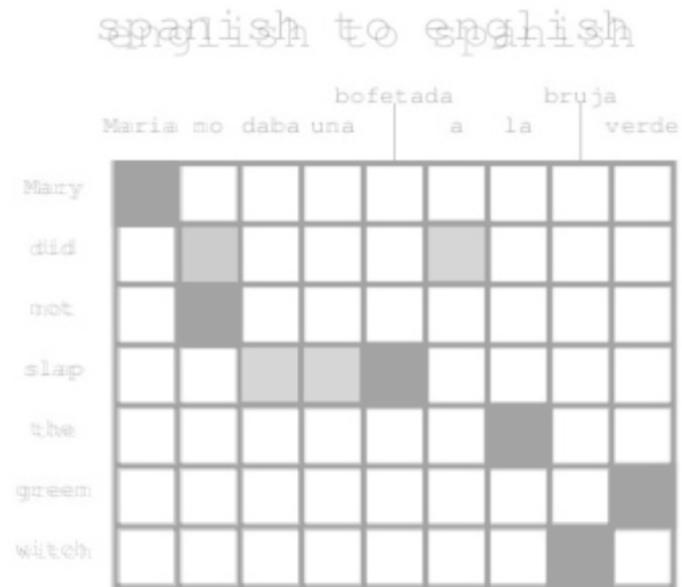
Mary	■								
did					■				
not		■							
slap			■	■	■	■			
the							■		
green									■
witch							■		

spanish to english

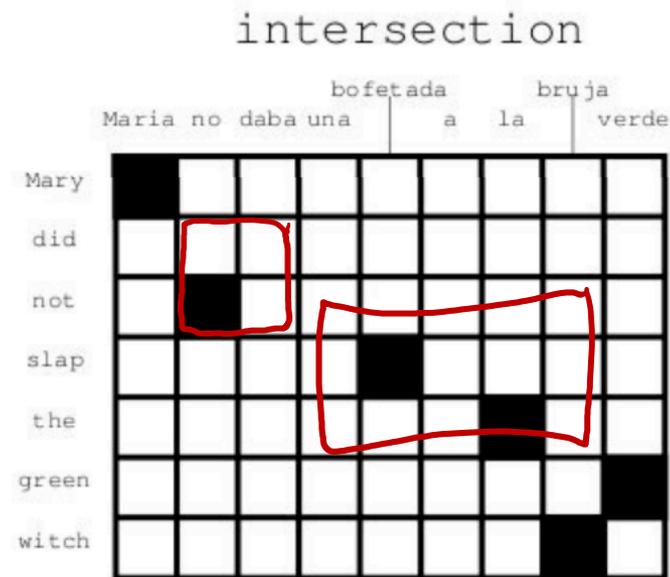
 bofetada bruja
Maria no daba una | a la | verde

Mary	■								
did		■							
not									
slap					■				
the							■		
green									■
witch							■		

Learning Phrasal Alignments

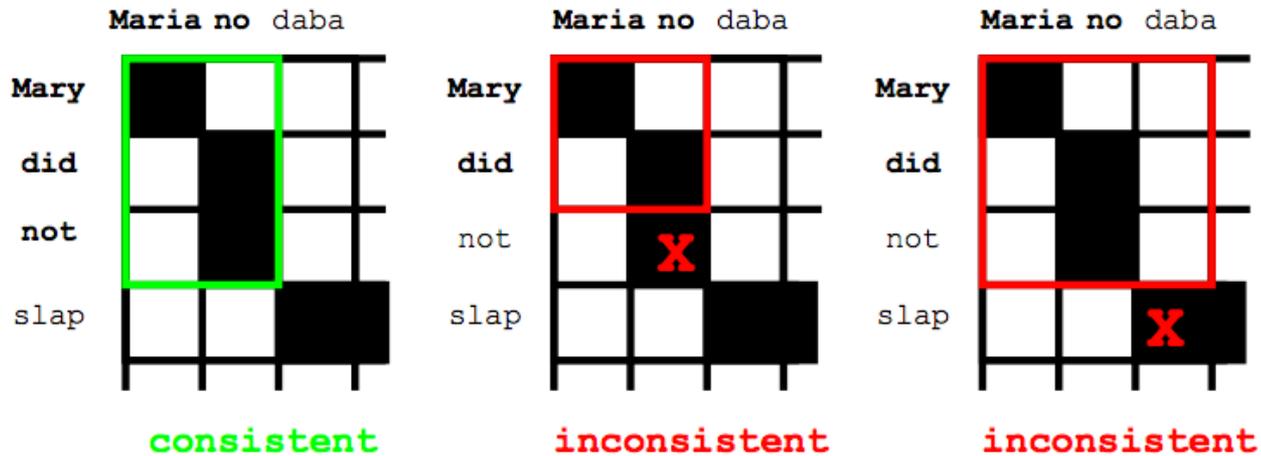


Learning Phrasal Alignments



Phrasal Alignments

Should contain all the alignment points for covered words



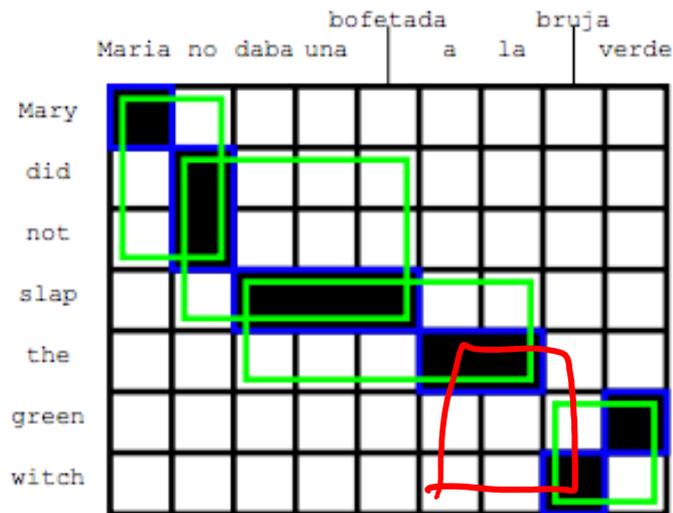
Learning Phrasal Alignments

intersection



(Maria, Mary), (no, did not),
 (slap, daba una bofetada),
 (a la, the), (bruja, witch),
 (verde, green)

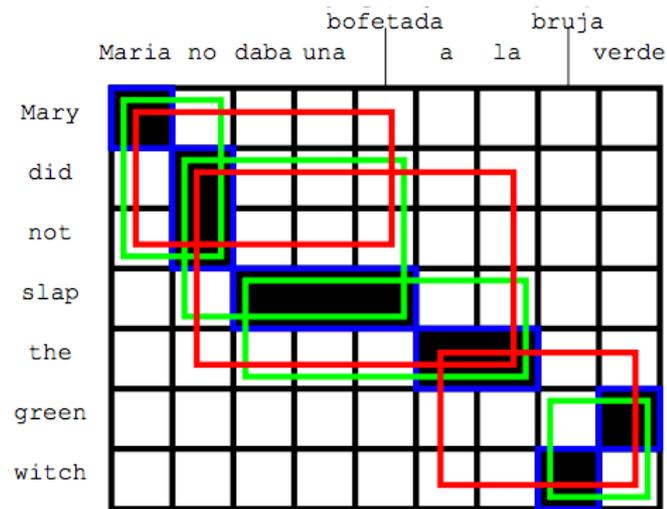
Learning Phrasal Alignments



(Maria, Mary), (no, did not),
(slap, daba una bofetada), (a la, the),
(bruja, witch), (verde, green)

(Maria no, Mary did not),
(no daba una bofetada, did not slap),
(daba una bofetada a la, slap the),
(bruja verde, green witch)

Learning Phrasal Alignments

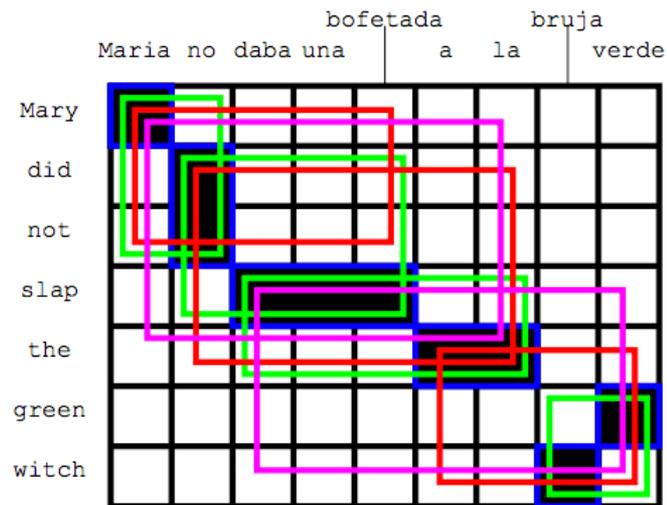


(Maria, Mary), (no, did not),
(slap, daba una bofetada), (a la, the),
(bruja, witch), (verde, green)

(Maria no, Mary did not),
(no daba una bofetada, did not slap),
(daba una bofetada a la, slap the),
(bruja verde, green witch)

(Maria no daba una bofetada, Mary did not slap),
(no daba una bofetada a la, did not slap the),
(a la bruja verde, the green witch)

Learning Phrasal Alignments



(Maria, Mary), (no, did not),
(slap, daba una bofetada), (a la, the),
(bruja, witch), (verde, green)

(Maria no, Mary did not),
(no daba una bofetada, did not slap),
(daba una bofetada a la, slap the),
(bruja verde, green witch)

(Maria no daba una bofetada, Mary did not slap),
(no daba una bofetada a la, did not slap the),
(a la bruja verde, the green witch)

(Maria no daba una bofetada a la, Mary did not slap the),
(daba una bofetada a la bruja verde, slap the green witch)

Phrase Translation Scores

$$P(f_p | e_p) = \frac{\text{count}(f_p, e_p)}{\text{count}(e_p)}$$

$$\mathcal{d} = \{(f_p, e_p, g_p)\}$$

Phrases for a Sentence

1 2 3 4 5 6 7 f
wir müssen auch diese kritik ernst nehmen — f

1 ↙

(wir müssen, we must)
(wir müssen auch, we must also)
(ernst, seriously)
....



$P \Rightarrow p = (s_i, e)$

(1, 2, we must)
(1, 3, we must also)
(6, 6, seriously)
⋮

Derivations for a Sentence

$$y = P_1 \dots P_L \quad \curvearrowright \quad P_i(s, t, e) \Rightarrow g_i$$

$$y = \begin{array}{l} (1, 3, \text{we must also}) (7, 7, \text{take}) \\ (4, 5, \text{this criticism}) (6, 6, \text{seriously}) \end{array}$$

$$e(y) = \text{we must also take this criticism seriously}$$

Distortion Limits, d

$$|t_k + 1 - s_{k+1}| \leq \underline{d}$$

$$\underbrace{(1, 3, \dots)}_5 (7, 7, \dots) (4, 5, \dots) (6, 6, \dots)$$

Distortion Scores

$$D(y) = \sum_k D(t_k - S_{k+1})$$

$$D(x) = \begin{cases} |x| > d & \infty \\ \text{o.w.} & \eta |x| \end{cases}$$

η ← $\alpha^{|x|}$

Scoring Derivations

$$S(y) = \log h(y) + \sum_k g(p_k) + D(y)$$

Language model Translation distortion score

$$\sum_i h(e_i | e_{i-2}, e_{i-1})$$

The Translation Problem

Given: \mathcal{L} , h , d , η

$f \longrightarrow e$

$\operatorname{argmax}_{y \in \mathcal{Y}(f)} s(y)$

A Secret of Statistical MT

$$P(e) P(f|e) P_a(f|e)^{13.7}$$

$$P(e)^{1.9} P(f|e) |e| P(e|f) \dots$$

$$s(y) = w \cdot \phi(y)$$

Outline

EM-Algorithm for Alignments

Phrase-Based MT

Decoding Algorithms

Syntax-Based MT

The Decoding Task

$$\operatorname{argmax}_y S(y)$$
$$\log h(y) + \sum_k g(p_k) + \sum_k d(p_k, p_{k+1})$$

$$S(h+p) = S(h) + S(p)$$

Monotonic Word 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 Word 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>by</u>			
	<u>no</u>							

Monotonic Word 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>by</u>			
	<u>no</u>							

Mary

Monotonic Word 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>by</u>			
	<u>no</u>							

Mary did not
/ not
/ no

Monotonic Word 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>by</u>			
	<u>no</u>							

Mary did not give
 not
 no

Monotonic Word 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>by</u>			
	<u>no</u>							

Mary did not give a

Monotonic Word 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>by</u>			
	<u>no</u>							

Mary did not give a slap

Monotonic Word Decoding

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

Mary did not give a slap — to
by

Monotonic Word Decoding

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

Mary did not give a slap to the
by

Monotonic Word Decoding

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

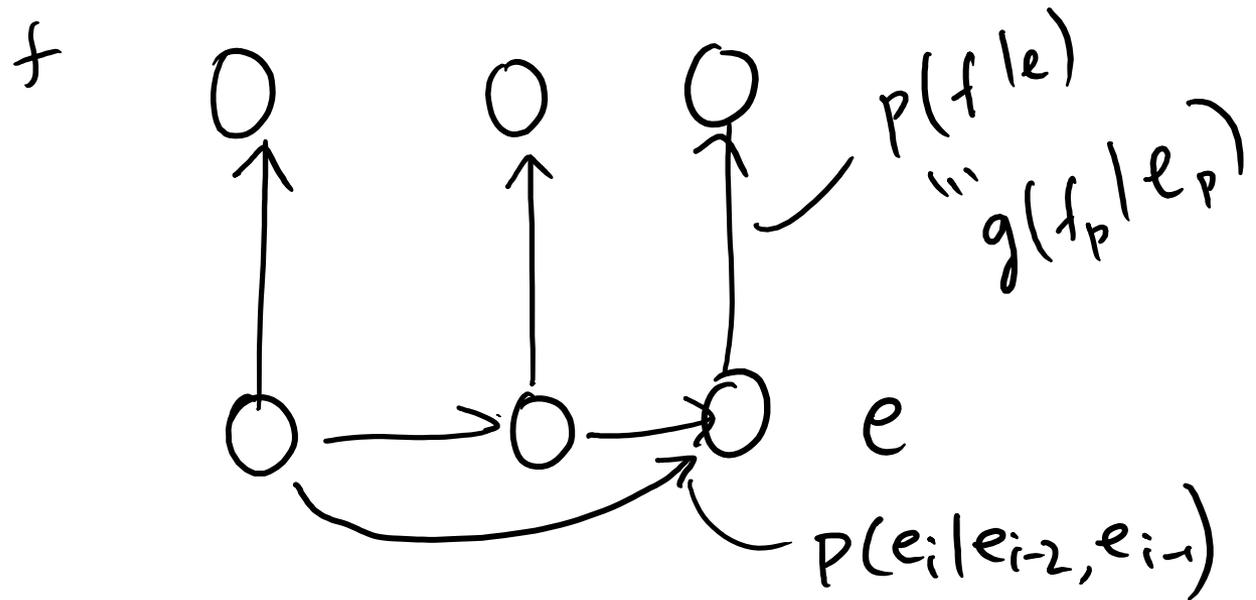
Mary did not give a slap to the witch

Monotonic Word 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>by</u>			
	<u>no</u>							

Mary did not give a slap to the witch green

Monotonic Word Decoding



Phrase Decoding: Stacks

state
E: english
F: how much of f
is covered
Cost: partial score

Phrase Decoding: Stacks

$E = ""$
 $F = ""$
 $C = 0.0$

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

Keep only B hypotheses

drop least scoring

$h \ E = "e_1 e_2"$

$F = * * \dots$

$C = \oplus$

$p = (s, t, e)$

\Downarrow

$h' \ E = "e_1 e_2 e"$ ^(s,t)

$F = * * * * \dots$

$C = C + s(p)$

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
Mary	<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>	

-

E = "Mary"

F = + - - - -

C = ~

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>	

(Mary)

-

E = "Mary not"

F = * * - - -

C

"Mary did not"

F = * * - - - -

C

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>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>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)