End-to-End Discriminative Training for Machine Translation

Percy Liang
Dan Klein

Alex Bouchard-Côté
Ben Taskar

UC Berkeley

Computer Science Division

Discriminative machine translation

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

Discriminative machine translation

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

Training examples

$\mathbf{x}_1$: mr barn crespo wishes to make a comment .
$\mathbf{y}_1$: m. barn crespo veut intervenir .

$\mathbf{x}_2$: i consider it very important to continue this work .
$\mathbf{y}_2$: à mon avis , il est très important de continuer dans cette voie-là .

$\mathbf{x}_3$: this all augurs well , and represents real progress .
$\mathbf{y}_3$: tout cela est de bon augure et constitue un réel progrès .

. . .

Discriminative machine translation

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

Training examples

Features

$\mathbf{x}_1$: mr barn crespo wishes to make a comment .
$\mathbf{y}_1$: m. barn crespo veut intervenir .

$\mathbf{x}_2$: i consider it very important to continue this work .
$\mathbf{y}_2$: à mon avis , il est très important de continuer dans cette voie-là .

$\mathbf{x}_3$: this all augurs well , and represents real progress .
$\mathbf{y}_3$: tout cela est de bon augure et constitue un réel progrès .

. . .

Discriminative machine translation

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

Training examples

$\mathbf{x}_1$: mr barn crespo wishes to make a comment .
$\mathbf{y}_1$: m. barn crespo veut intervenir .

$\mathbf{x}_2$: i consider it very important to continue this work .
$\mathbf{y}_2$: à mon avis , il est très important de continuer dans cette voie-là .

$\mathbf{x}_3$: this all augurs well , and represents real progress .
$\mathbf{y}_3$: tout cela est de bon augure et constitue un réel progrès .

. . .

Features

NN JJ $\rightarrow$ JJ NN 1

Discriminative machine translation

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

Training examples

$\mathbf{x}_1$: mr barn crespo wishes to make a comment .
$\mathbf{y}_1$: m. barn crespo veut intervenir .

$\mathbf{x}_2$: i consider it very important to continue this work .
$\mathbf{y}_2$: à mon avis , il est très important de continuer dans cette voie-là .

$\mathbf{x}_3$: this all augurs well , and represents real progress .
$\mathbf{y}_3$: tout cela est de bon augure et constitue un réel progrès .

. . .

Features

NN JJ $\rightarrow$ JJ NN	1
Lang. model	-4.253

Discriminative machine translation

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

Training examples

$\mathbf{x}_1$: mr barn crespo wishes to make a comment .
$\mathbf{y}_1$: m. barn crespo veut intervenir .

$\mathbf{x}_2$: i consider it very important to continue this work .
$\mathbf{y}_2$: à mon avis , il est très important de continuer dans cette voie-là .

$\mathbf{x}_3$: this all augurs well , and represents real progress .
$\mathbf{y}_3$: tout cela est de bon augure et constitue un réel progrès .

. . .

Features

NN JJ $\rightarrow$ JJ NN	1
Lang. model	-4.253

. . .

Discriminative machine translation

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

Training examples

$\mathbf{x}_1$: mr barn crespo wishes to make a comment .
$\mathbf{y}_1$: m. barn crespo veut intervenir .

$\mathbf{x}_2$: i consider it very important to continue this work .
$\mathbf{y}_2$: à mon avis , il est très important de continuer dans cette voie-là .

$\mathbf{x}_3$: this all augurs well , and represents real progress .
$\mathbf{y}_3$: tout cela est de bon augure et constitue un réel progrès .

. . .

Features

		$\mathbf{w}$
NN JJ $\rightarrow$ JJ NN	1	2.3
Lang. model	-4.253	1.7
. . .		

Discriminative machine translation

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

Training examples

$\mathbf{x}_1$: mr barn crespo wishes to make a comment .
$\mathbf{y}_1$: m. barn crespo veut intervenir .

$\mathbf{x}_2$: i consider it very important to continue this work .
$\mathbf{y}_2$: à mon avis , il est très important de continuer dans cette voie-là .

$\mathbf{x}_3$: this all augurs well , and represents real progress .
$\mathbf{y}_3$: tout cela est de bon augure et constitue un réel progrès .

. . .

That's it.

Features

		$\mathbf{w}$
NN JJ $\rightarrow$ JJ NN	1	2.3
Lang. model	-4.253	1.7
. . .		

Discriminative machine translation

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

Training examples

$\mathbf{x}_1$: mr barn crespo wishes to make a comment .
$\mathbf{y}_1$: m. barn crespo veut intervenir .

$\mathbf{x}_2$: i consider it very important to continue this work .
$\mathbf{y}_2$: à mon avis , il est très important de continuer dans cette voie-là .

$\mathbf{x}_3$: this all augurs well , and represents real progress .
$\mathbf{y}_3$: tout cela est de bon augure et constitue un réel progrès .

. . .

Features

		$\mathbf{w}$
NN JJ $\rightarrow$ JJ NN	1	2.3
Lang. model	-4.253	1.7
. . .		

That's it. Well, actually. . .

Why is this hard?

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

Why is this hard?

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

- Correct output is ill-defined

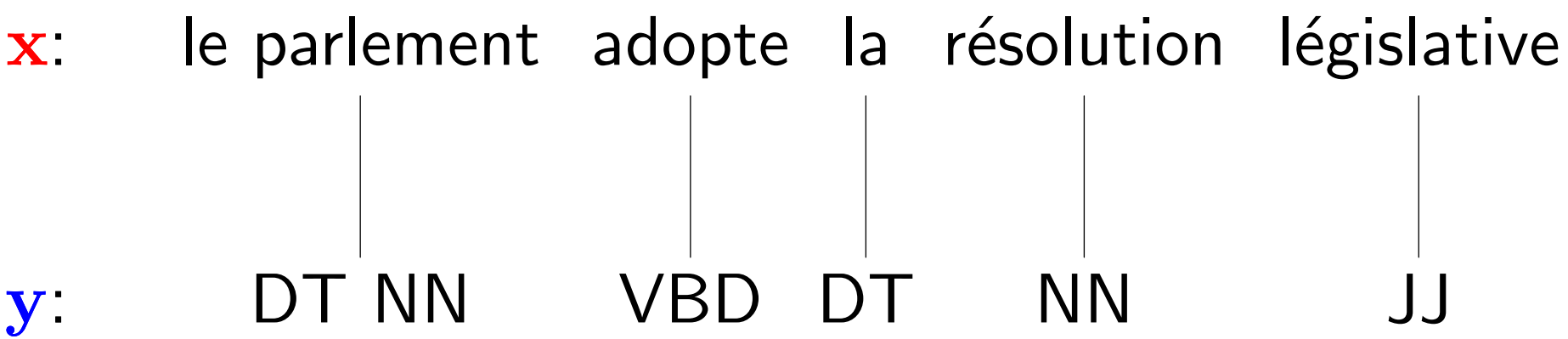
Why is this hard?

$\mathbf{x}$: le parlement adopte la résolution législative

$\mathbf{y}$: parliament adopted the legislative resolution

- Correct output is ill-defined
- Correspondences are missing

Why is this hard?



- Correct output is ill-defined
- Correspondences are missing

Why is this hard?

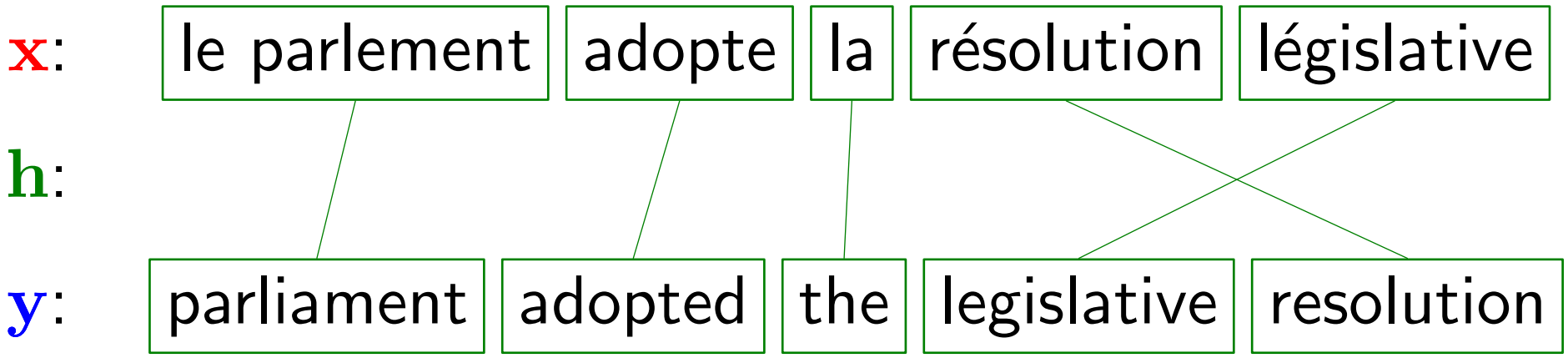
x: le parlement adopte la résolution législative

???

y: parliament adopted the legislative resolution

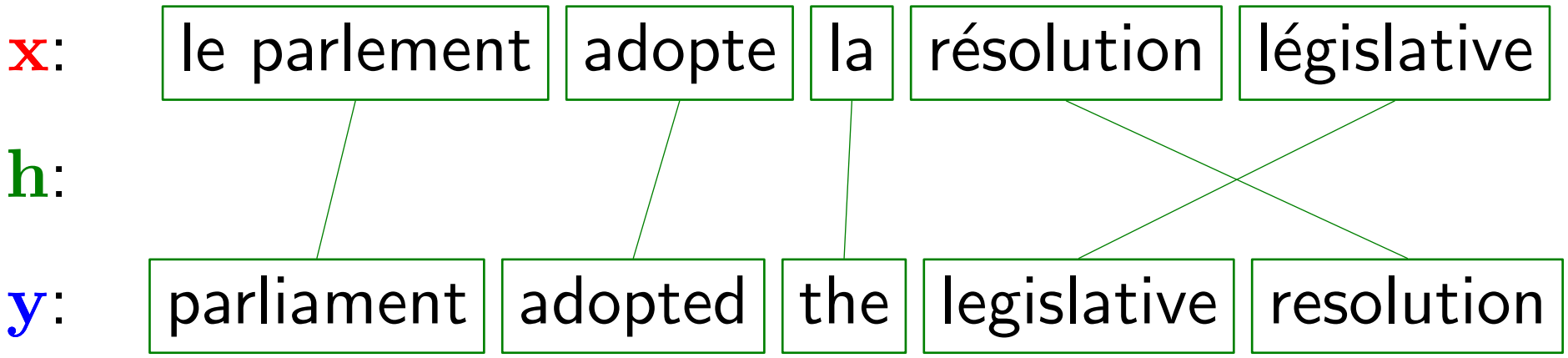
- Correct output is ill-defined
- Correspondences are missing

Why is this hard?



- Correct output is ill-defined
- Correspondences are missing

Why is this hard?



- Correct output is ill-defined
- Correspondences are missing
- Hidden correspondence is abused

Why is this hard?

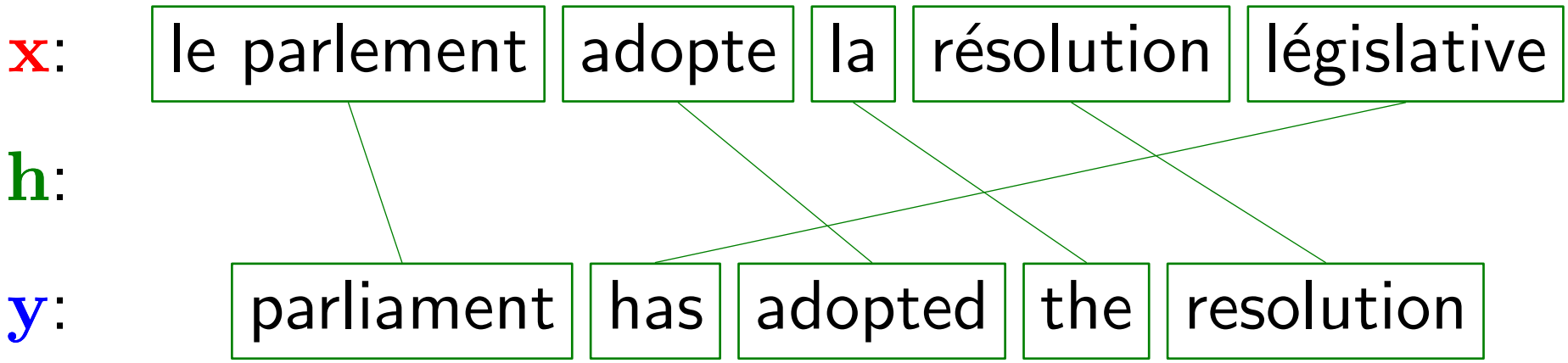
x: le parlement adopte la résolution législative

h:

y: parliament has adopted the resolution

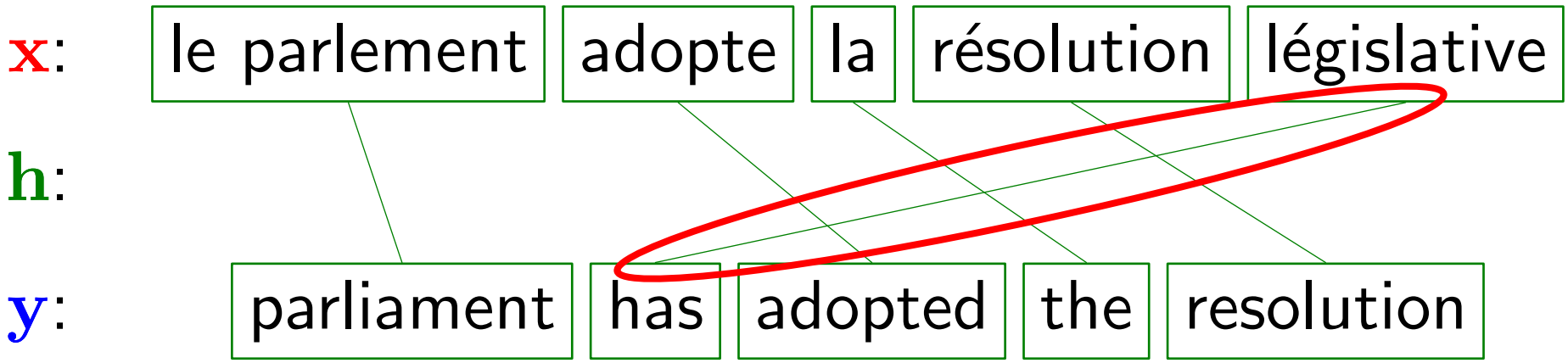
- Correct output is ill-defined
- Correspondences are missing
- Hidden correspondence is abused

Why is this hard?



- Correct output is ill-defined
- Correspondences are missing
- Hidden correspondence is abused

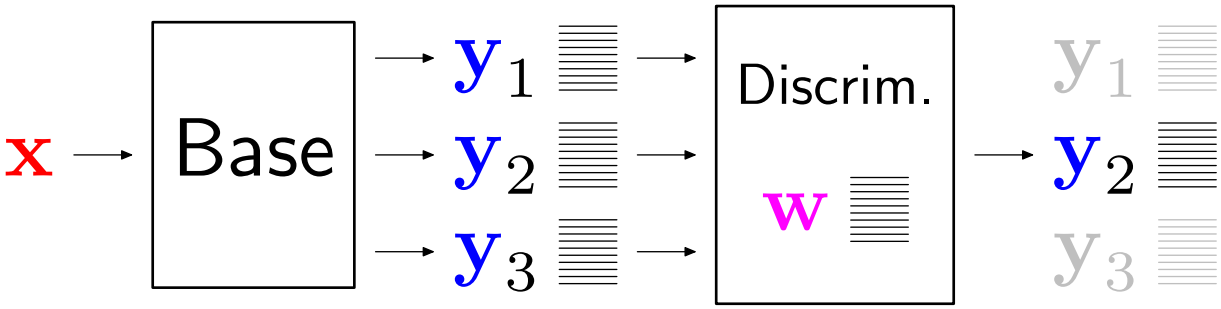
Why is this hard?



- Correct output is ill-defined
- Correspondences are missing
- Hidden correspondence is abused

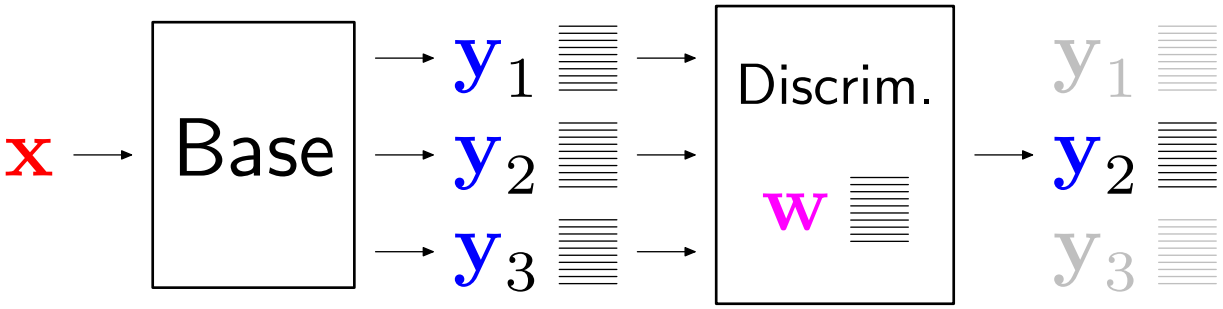
Discriminative approaches

Reranking [Shen, et al. '04; Och, et al. '04]

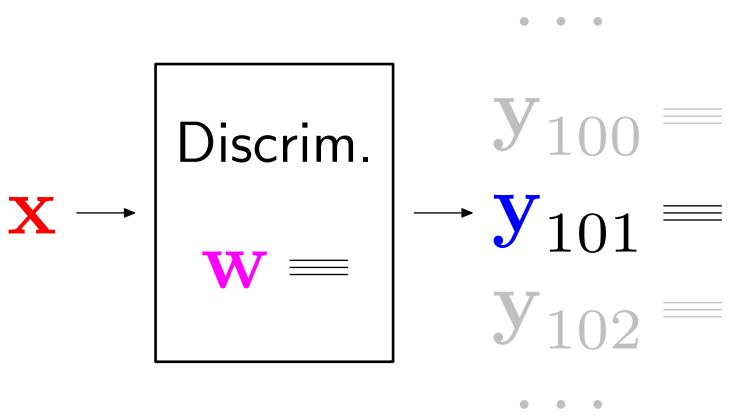


Discriminative approaches

Reranking [Shen, et al. '04; Och, et al. '04]

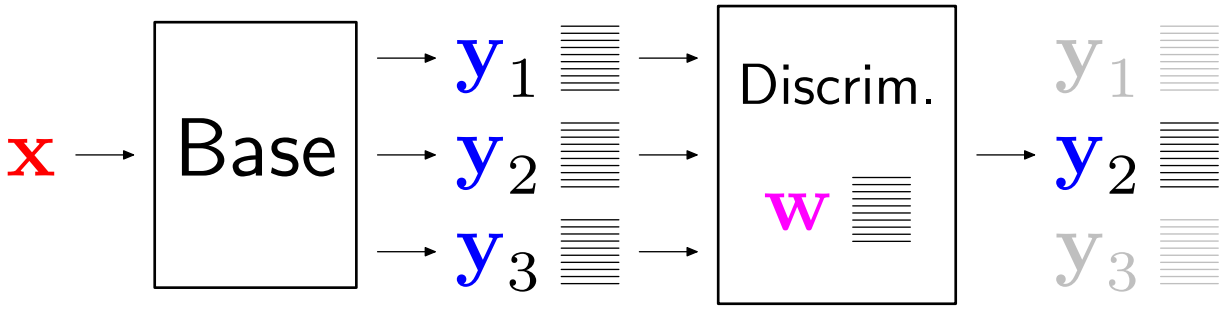


MERT [Och '03]

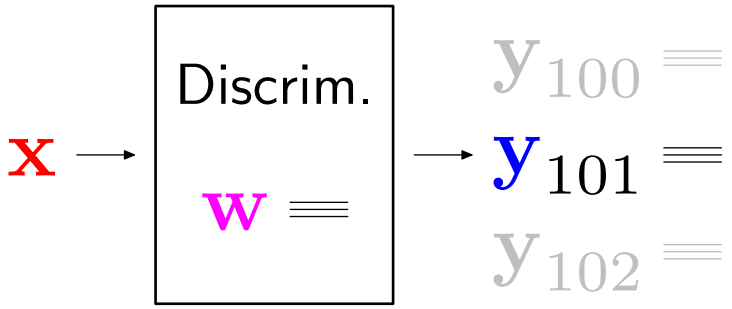


Discriminative approaches

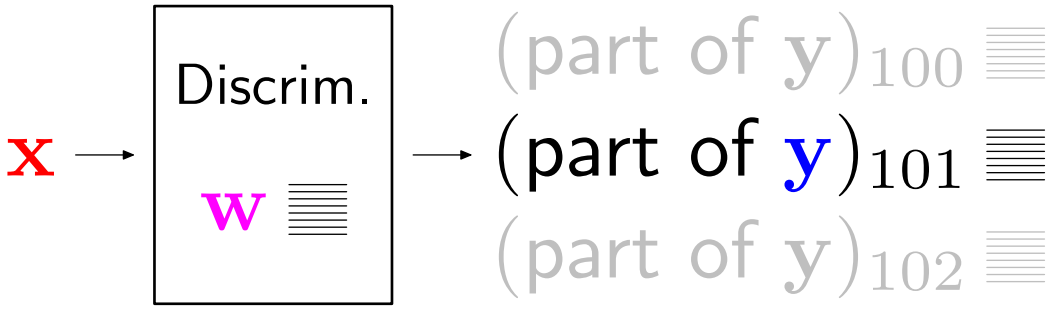
Reranking [Shen, et al. '04; Och, et al. '04]



MERT [Och '03]

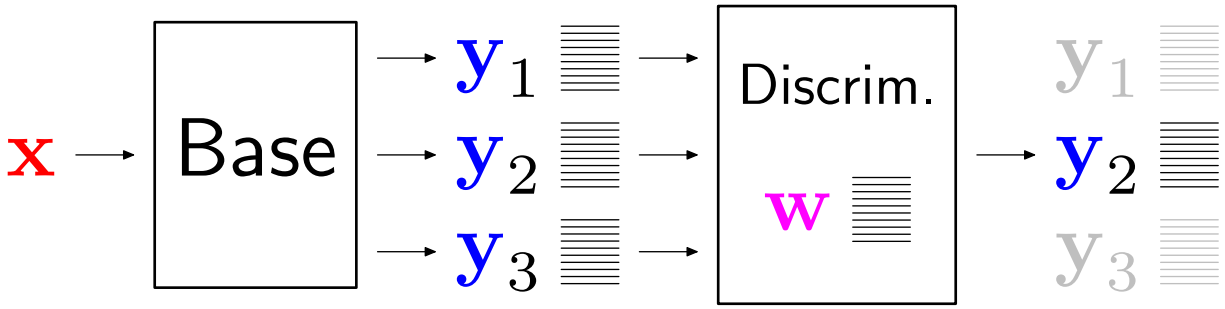


Local [Tillmann, Zhang '05]

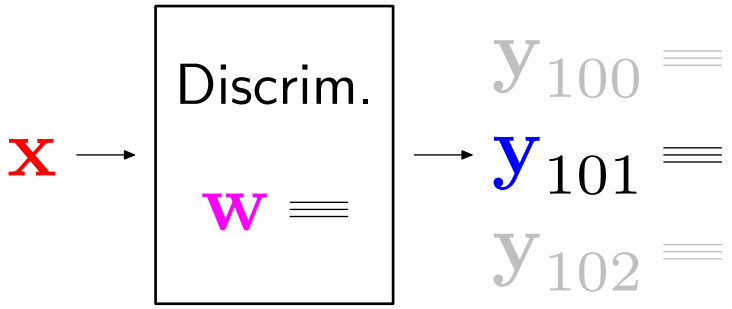


Discriminative approaches

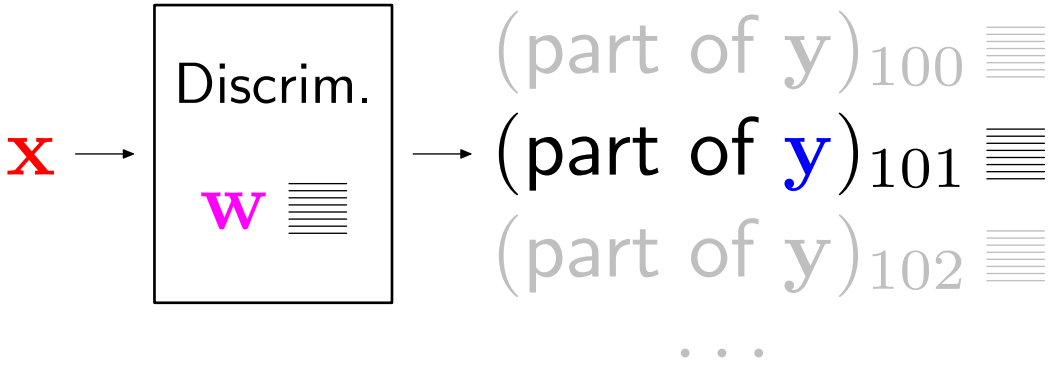
Reranking [Shen, et al. '04; Och, et al. '04]



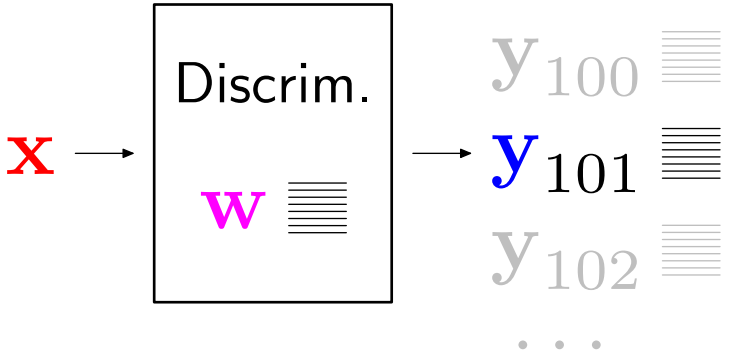
MERT [Och '03]



Local [Tillmann, Zhang '05]



Our end-to-end approach



Experimental setup

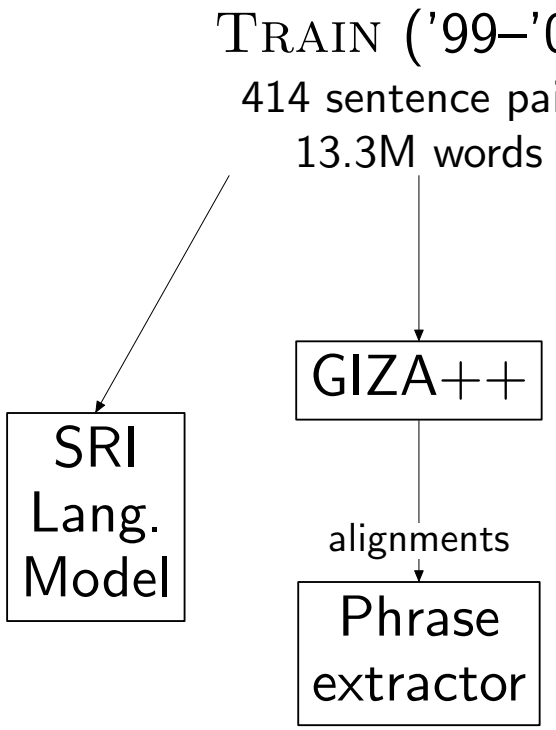
TRAIN ('99–'01)
414 sentence pairs
13.3M words

French-English Europarl corpus

DEV.5–15 ('02)
first 1K sentence pairs
10.4K words

TEST.5–15 ('03)
first 1K sentence pairs
10.8K words

Experimental setup

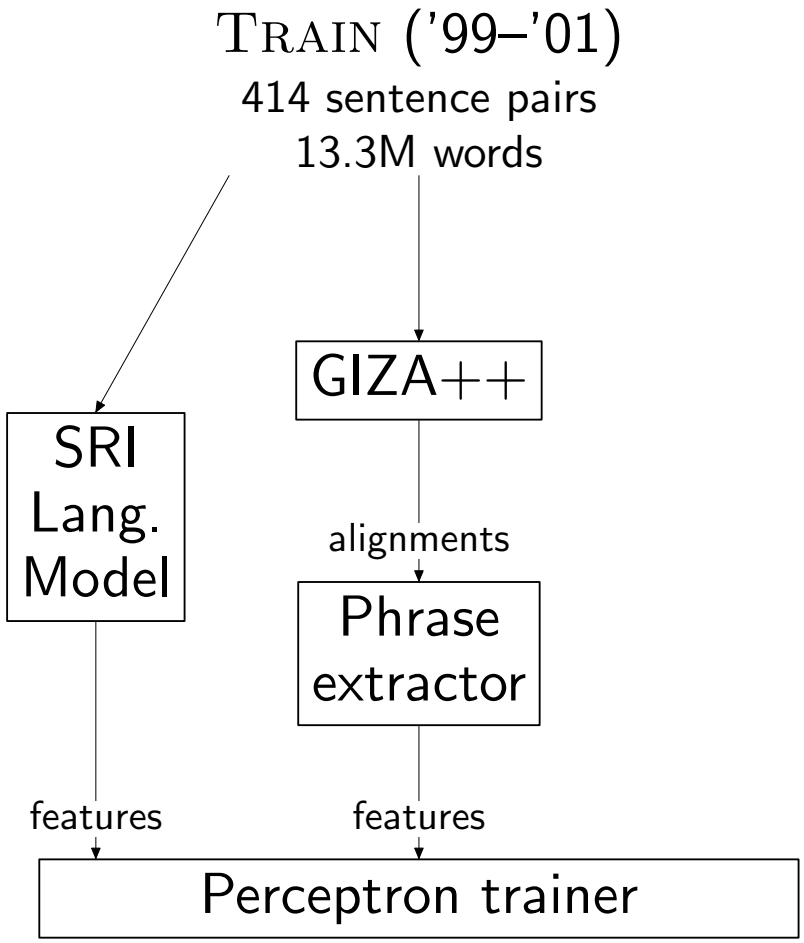


French-English Europarl corpus

DEV.5–15 ('02)
first 1K sentence pairs
10.4K words

TEST.5–15 ('03)
first 1K sentence pairs
10.8K words

Experimental setup

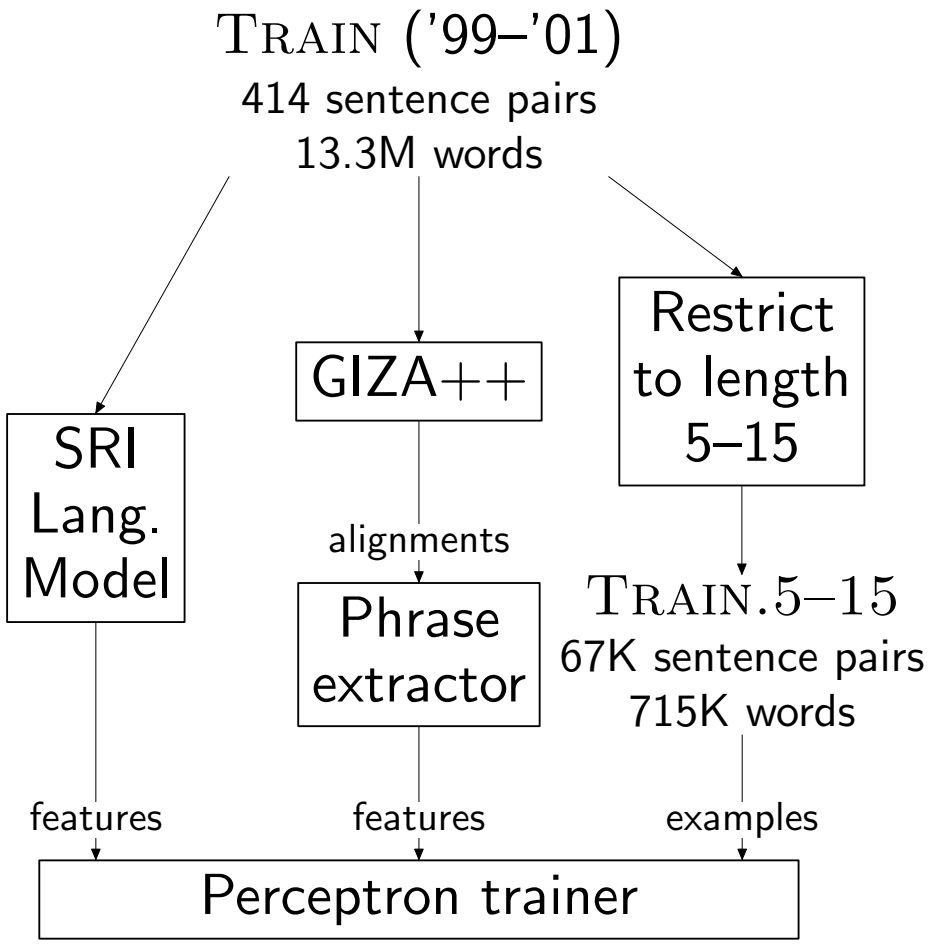


French-English Europarl corpus

DEV.5–15 ('02)
first 1K sentence pairs
10.4K words

TEST.5–15 ('03)
first 1K sentence pairs
10.8K words

Experimental setup

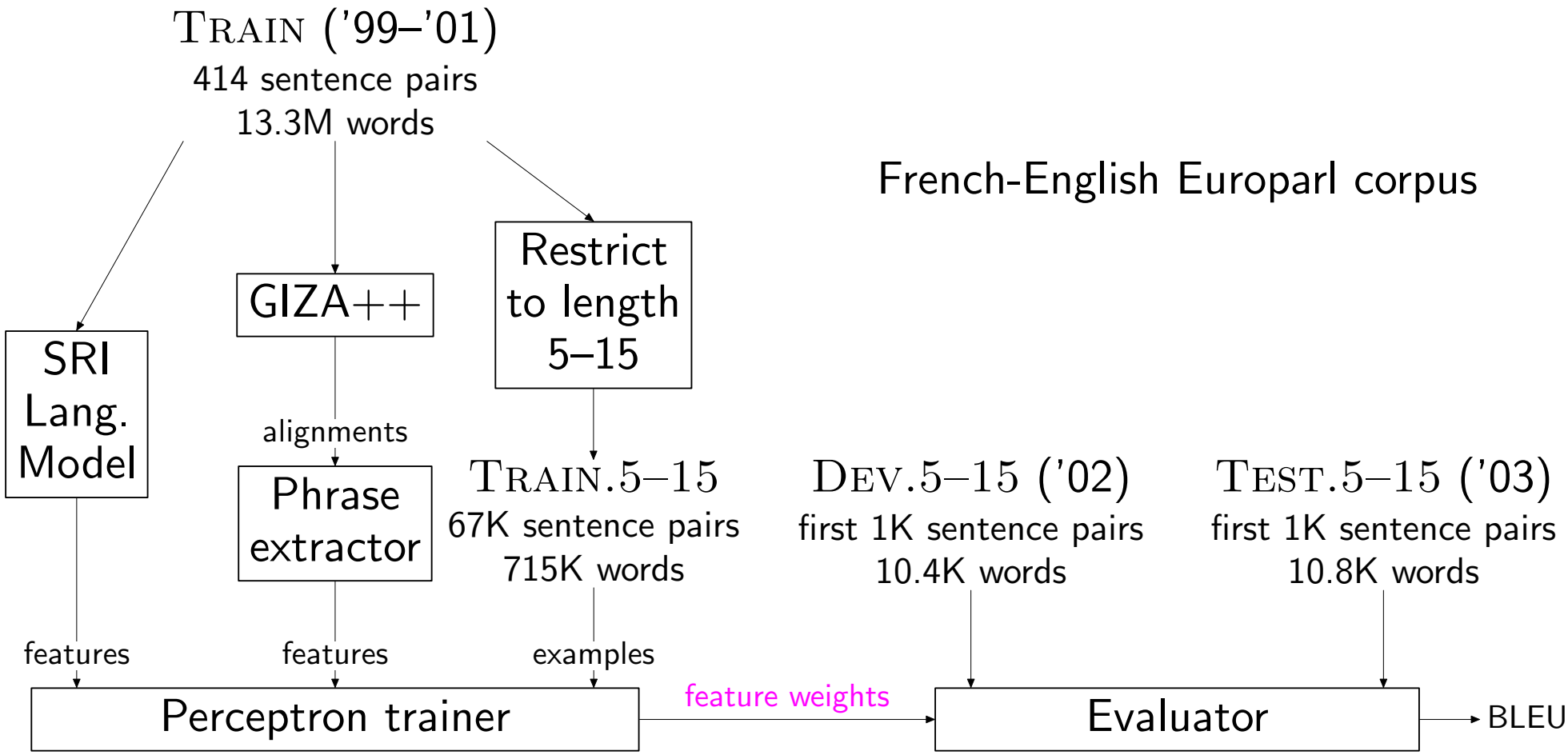


French-English Europarl corpus

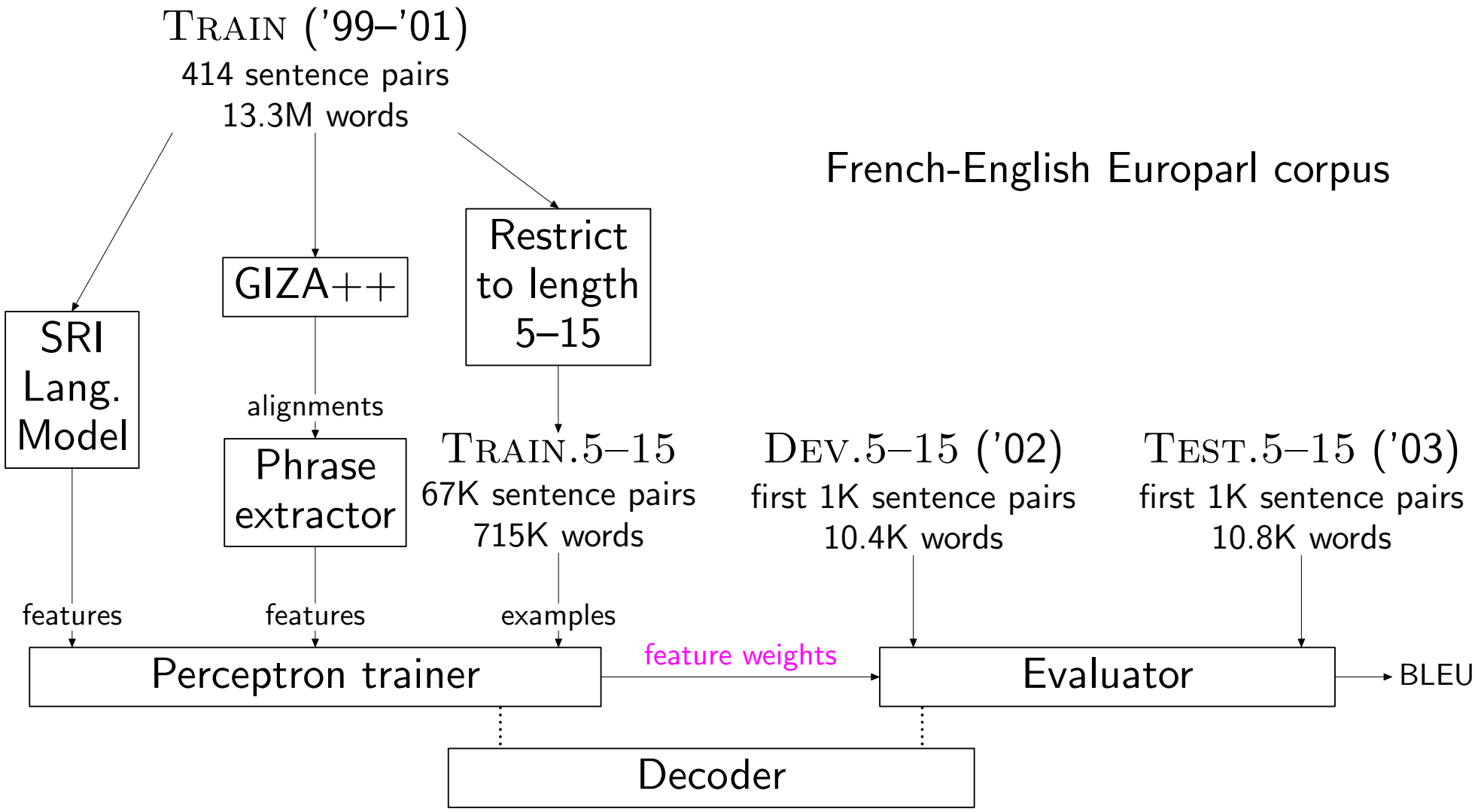
DEV.5–15 ('02)
first 1K sentence pairs
10.4K words

TEST.5–15 ('03)
first 1K sentence pairs
10.8K words

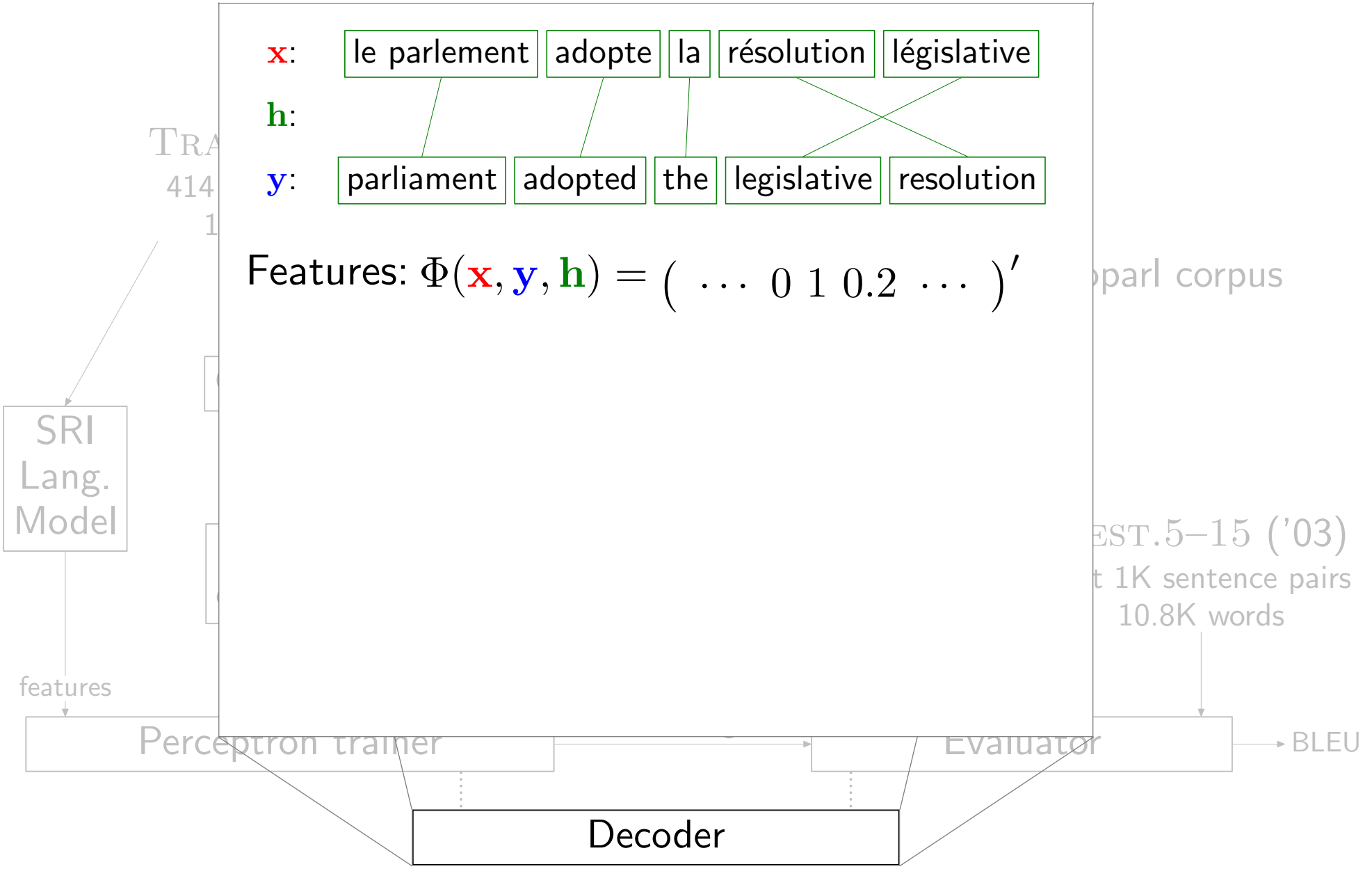
Experimental setup



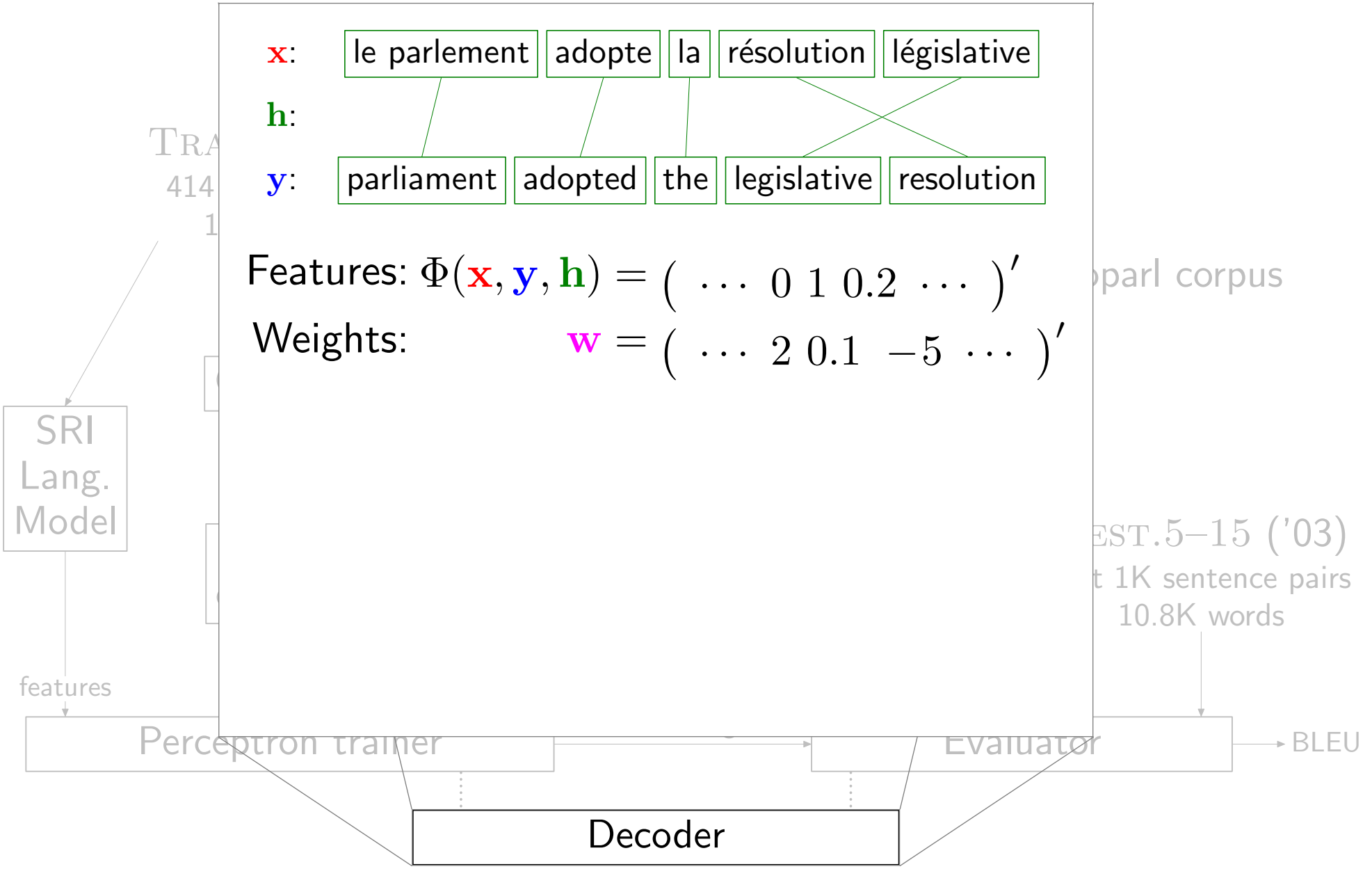
Experimental setup



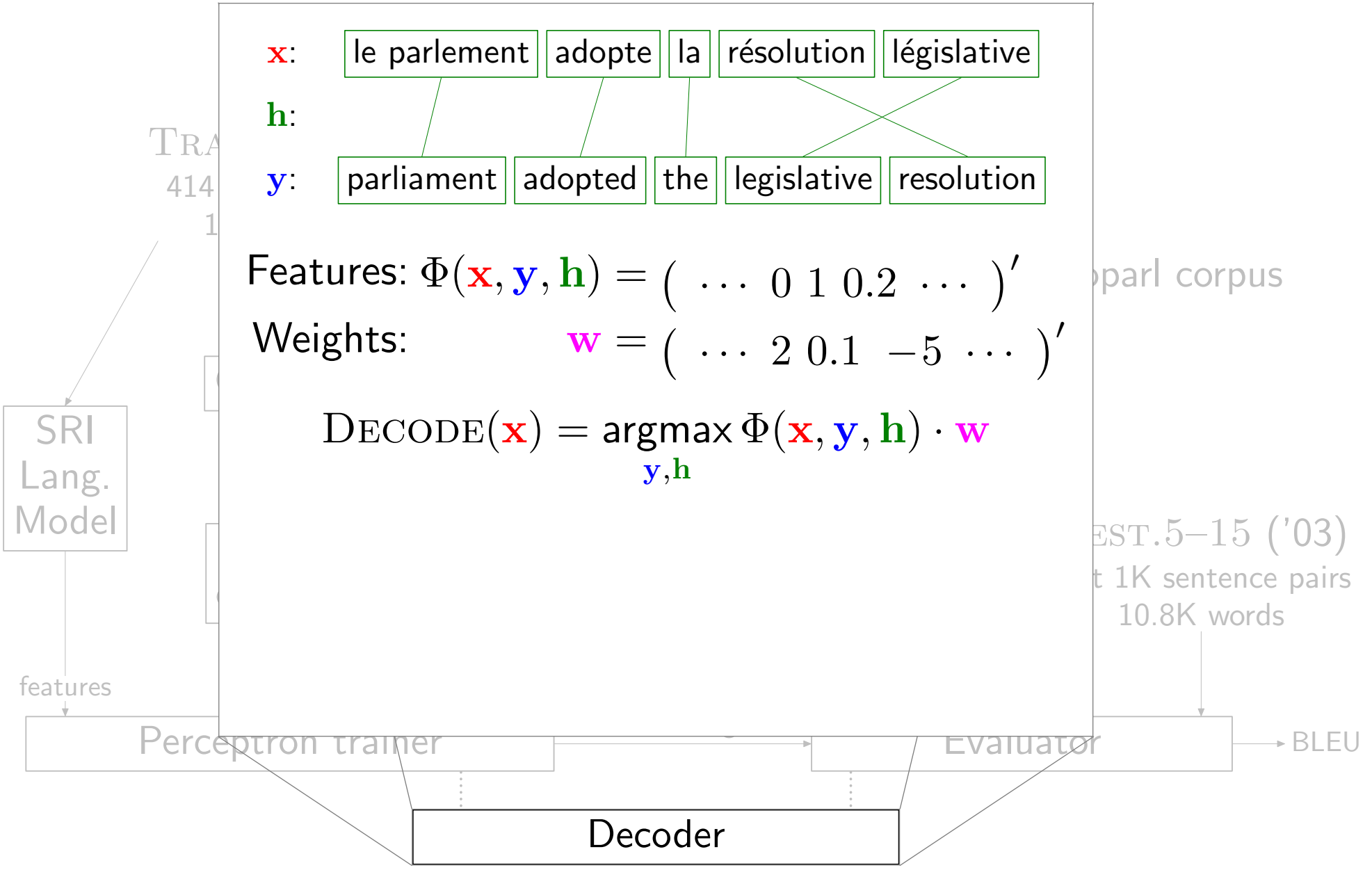
Experimental setup



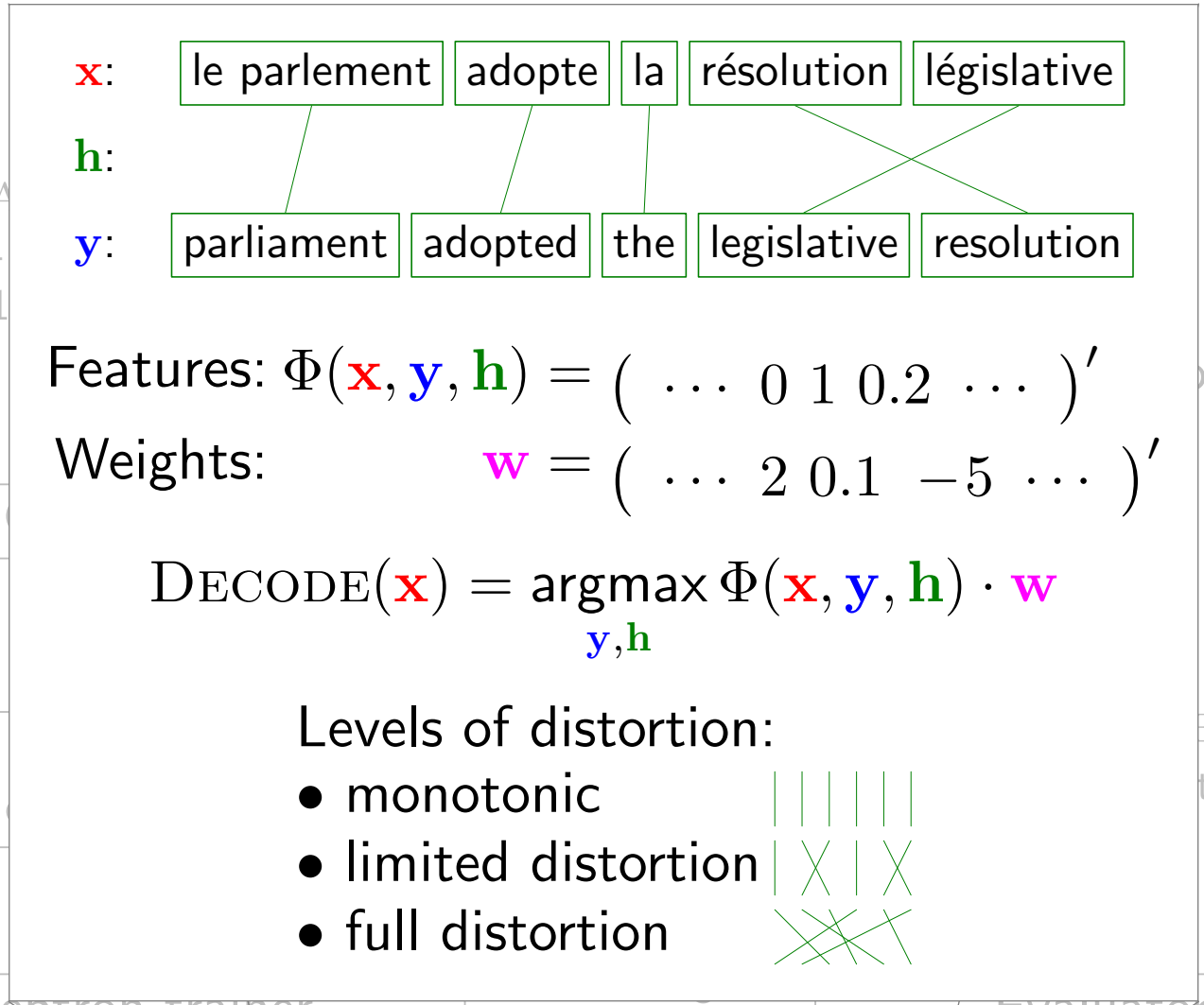
Experimental setup



Experimental setup



Experimental setup



Features: $\Phi(\mathbf{x}, \mathbf{y}, \mathbf{h}) = \begin{pmatrix} \cdots & 0 & 1 & 0.2 & \cdots \end{pmatrix}'$

Weights: $\mathbf{w} = \begin{pmatrix} \cdots & 2 & 0.1 & -5 & \cdots \end{pmatrix}'$

$$\textsc{Decode}(\mathbf{x}) = \underset{\mathbf{y},\mathbf{h}}{\text{argmax}}\, \Phi(\mathbf{x}, \mathbf{y}, \mathbf{h}) \cdot \mathbf{w}$$

Levels of distortion:

- monotonic
- limited distortion
- full distortion

Perceptron training

For each training example $(\mathbf{x}, \mathbf{y})$: [Collins '02]

$$\mathbf{w} \leftarrow \mathbf{w} \begin{array}{ll} +\Phi(\mathbf{x}, \mathbf{y}_\mathsf{t}) & \mid \mathbf{y}_\mathsf{t} = \mathbf{y} \\ -\Phi(\mathbf{x}, \mathbf{y}_\mathsf{p}) & \mid \mathbf{y}_\mathsf{p} = \textsc{Decode}(\mathbf{x}) \end{array}$$

Perceptron training

For each training example $(\mathbf{x}, \mathbf{y})$: [Collins '02]

$$\begin{aligned} \mathbf{w} \leftarrow \mathbf{w} \; & +\Phi(\mathbf{x}, \mathbf{y}_t) \\ & -\Phi(\mathbf{x}, \mathbf{y}_p) \end{aligned} \quad \Big| \quad \begin{aligned} \mathbf{y}_t &= \mathbf{y} \\ \mathbf{y}_p &= \textsc{Decode}(\mathbf{x}) \end{aligned}$$

$$\begin{aligned} \mathbf{w} \leftarrow \mathbf{w} \; & +\Phi(\mathbf{x}, \mathbf{y}_t, \mathbf{h}_t) \\ & -\Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p) \end{aligned} \quad \Big| \quad \begin{aligned} \mathbf{y}_t, \mathbf{h}_t &= \mathbf{???} \\ \mathbf{y}_p, \mathbf{h}_p &= \textsc{Decode}(\mathbf{x}) \end{aligned}$$

Perceptron training

For each training example $(\mathbf{x}, \mathbf{y})$: [Collins '02]

$$\begin{aligned} \mathbf{w} \leftarrow \mathbf{w} \; & +\Phi(\mathbf{x}, \mathbf{y}_t) \\ & -\Phi(\mathbf{x}, \mathbf{y}_p) \end{aligned} \quad \Big| \quad \begin{aligned} \mathbf{y}_t &= \mathbf{y} \\ \mathbf{y}_p &= \textsc{Decode}(\mathbf{x}) \end{aligned}$$

$$\begin{aligned} \mathbf{w} \leftarrow \mathbf{w} \; & +\Phi(\mathbf{x}, \mathbf{y}_t, \mathbf{h}_t) \\ & -\Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p) \end{aligned} \quad \Big| \quad \begin{aligned} \mathbf{y}_t, \mathbf{h}_t &= \mathbf{???} \\ \mathbf{y}_p, \mathbf{h}_p &= \textsc{Decode}(\mathbf{x}) \end{aligned}$$

How to choose $\mathbf{y}_t, \mathbf{h}_t$?

There are several choices and the choice does matter

Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure

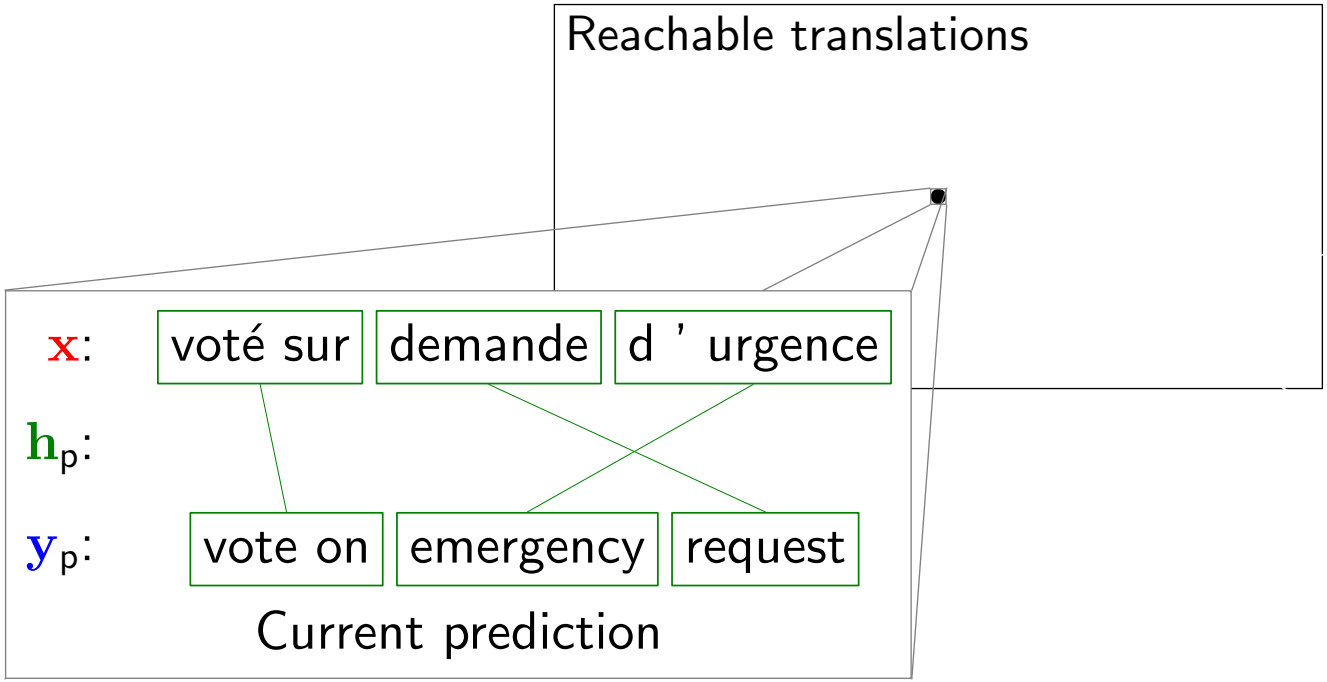
Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure



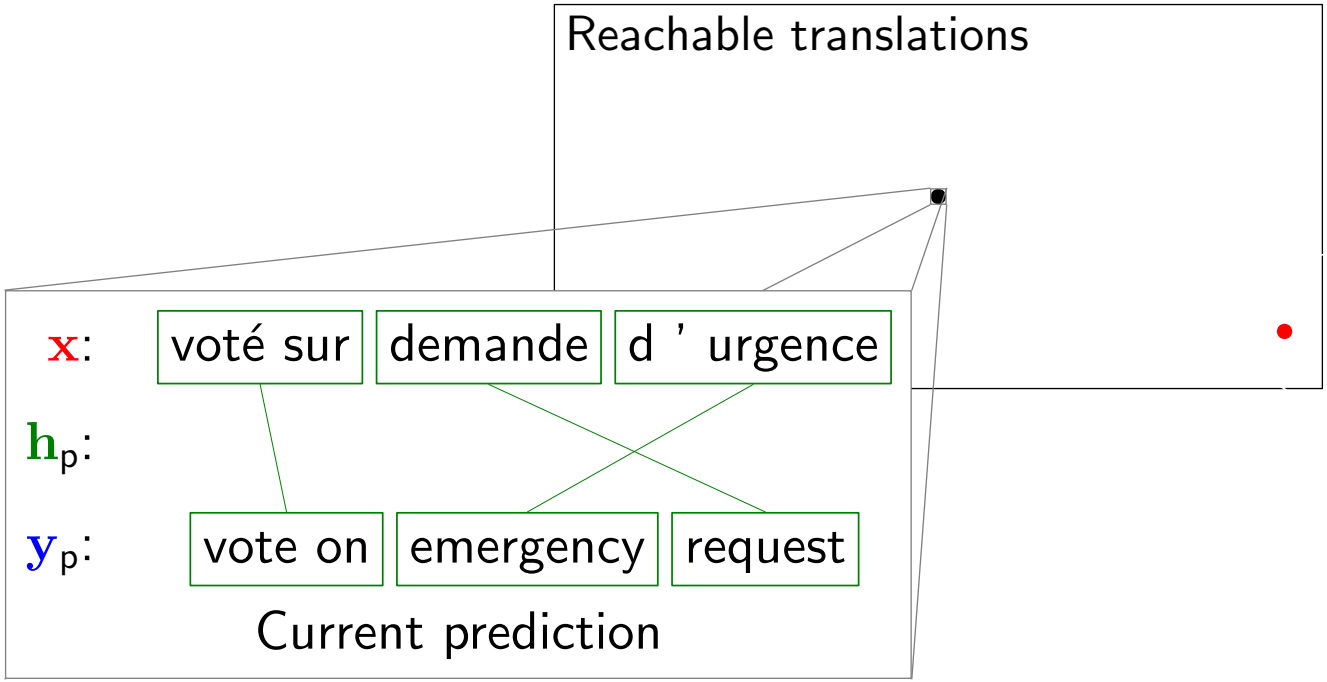
Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure



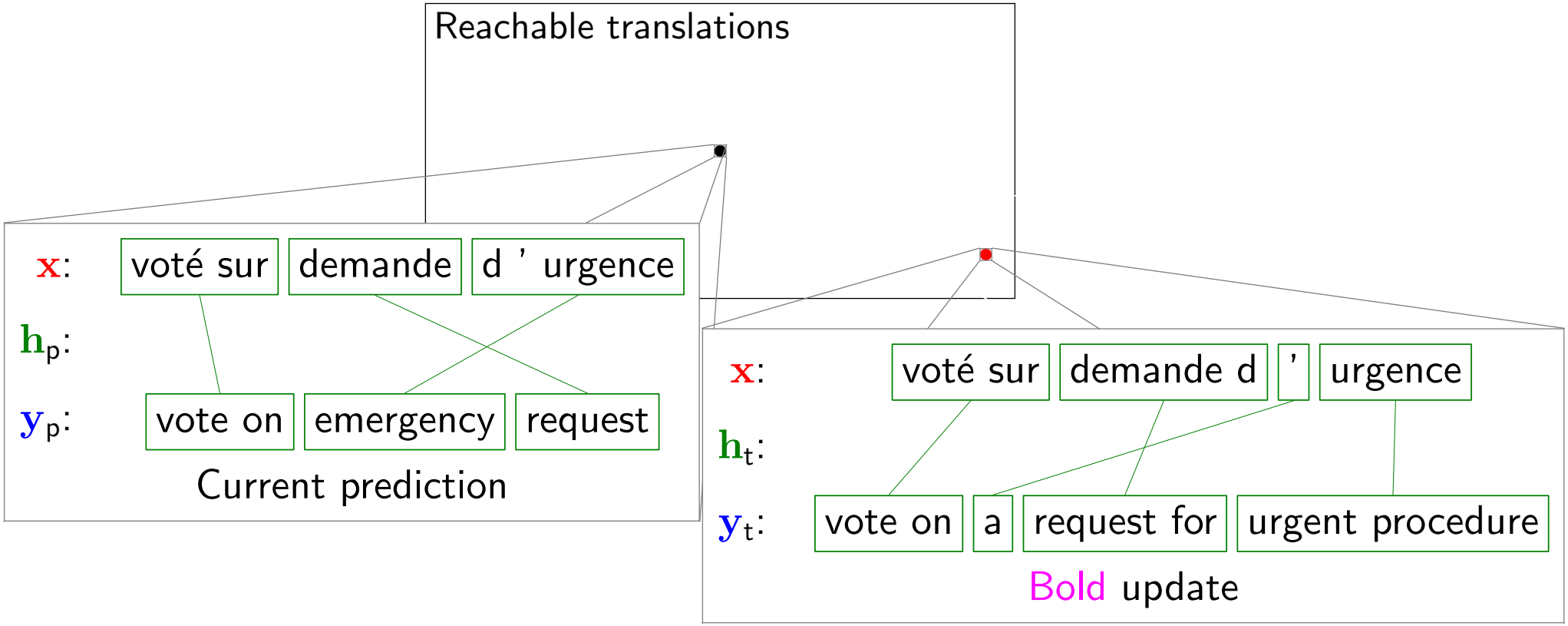
Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure



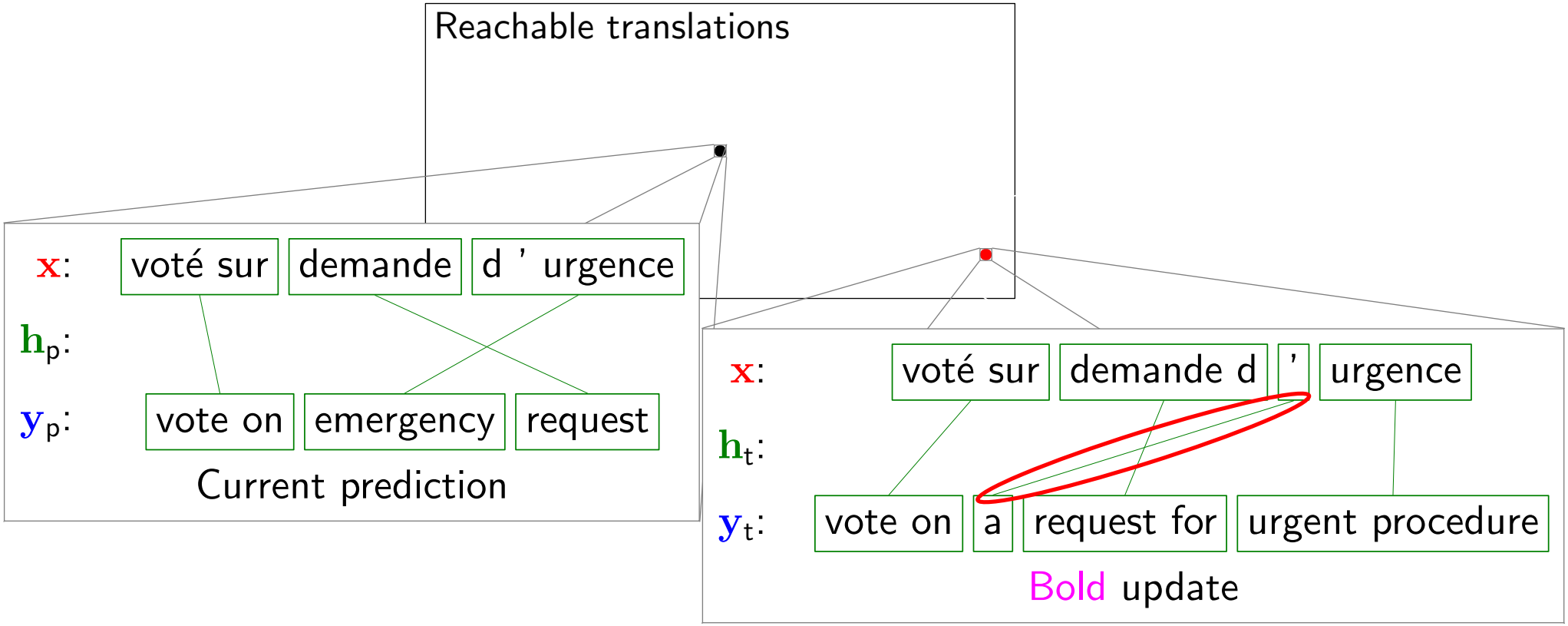
Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure



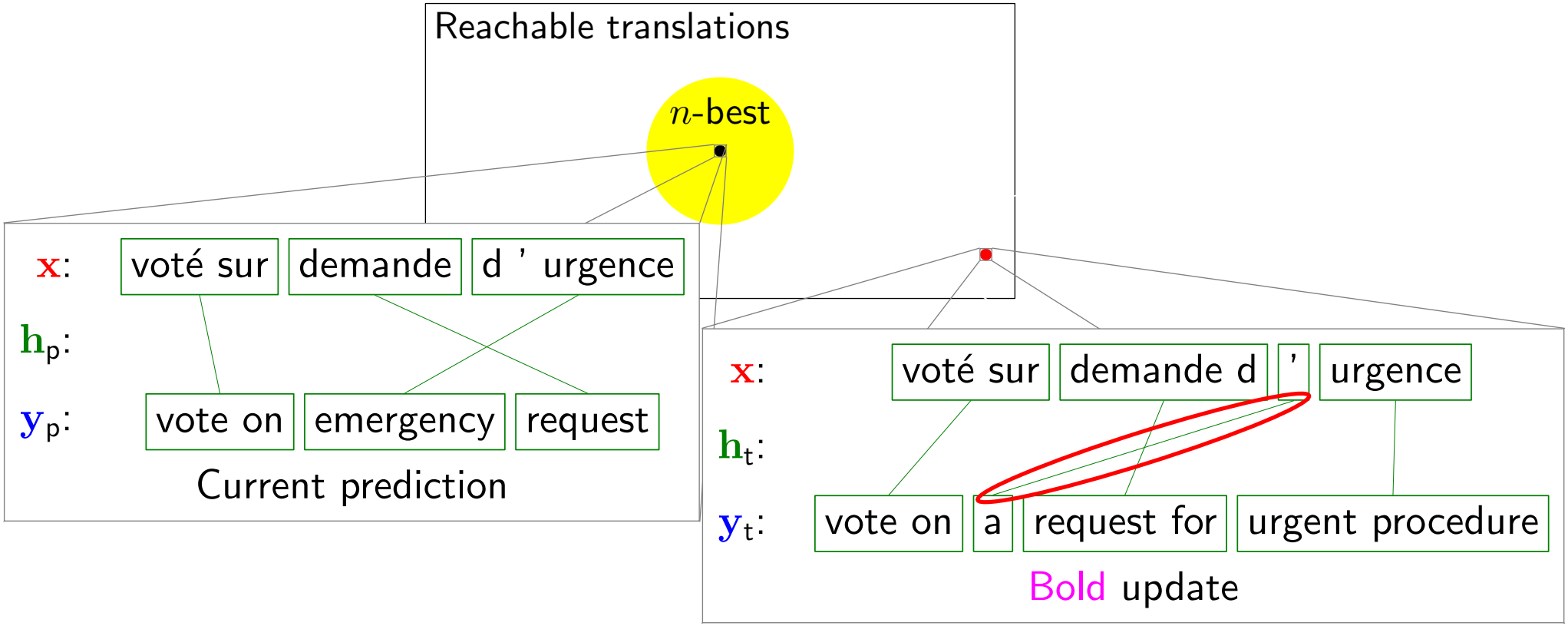
Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure



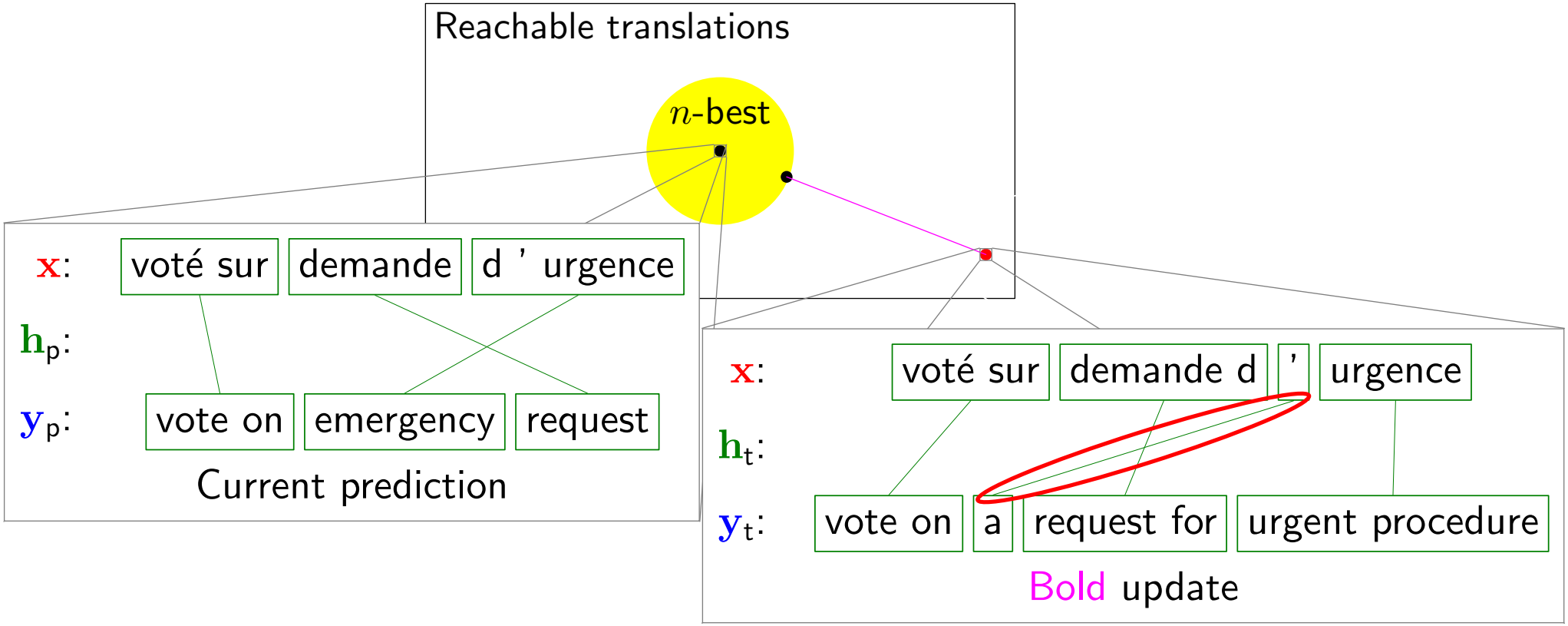
Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure



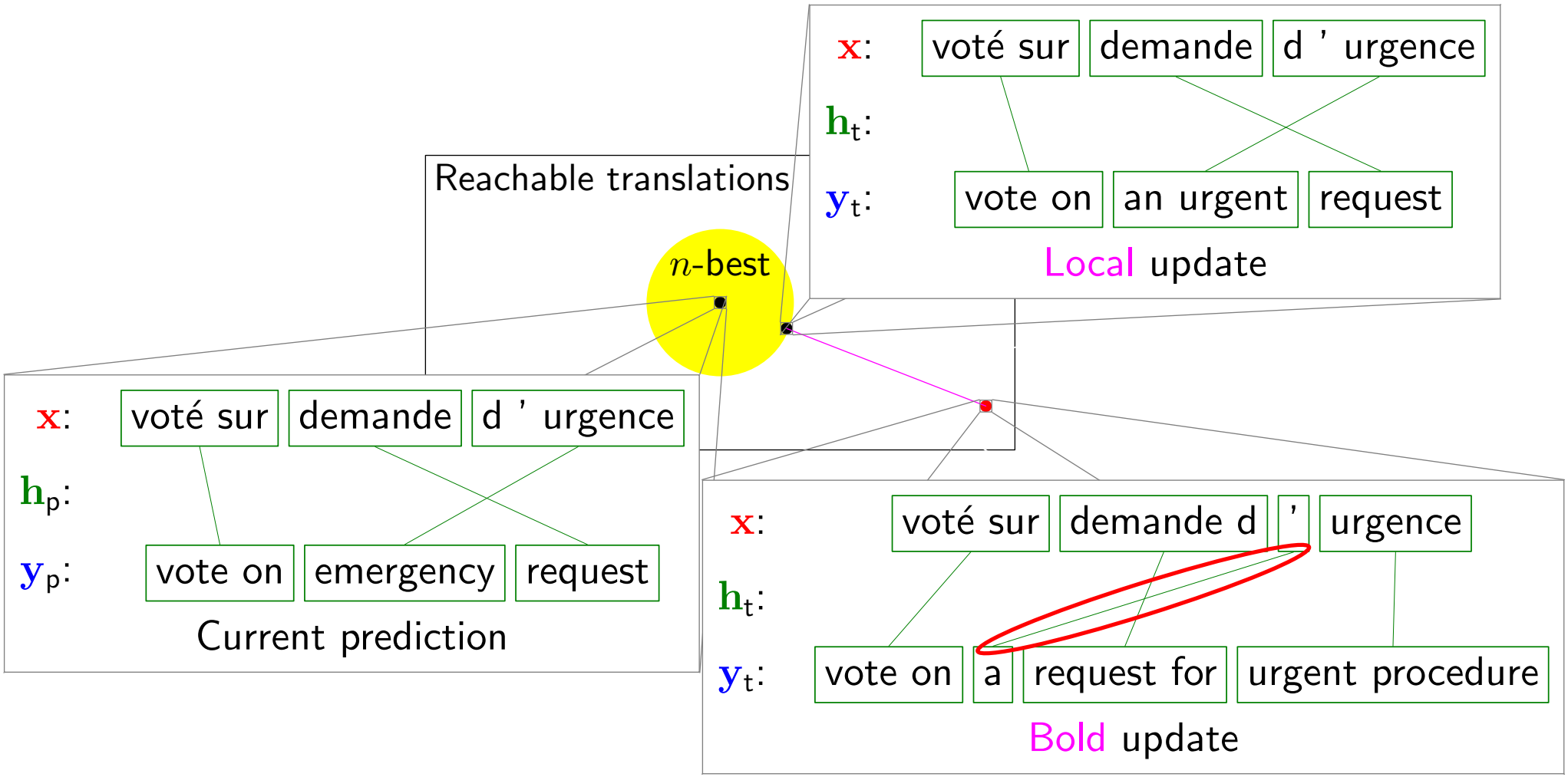
Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure



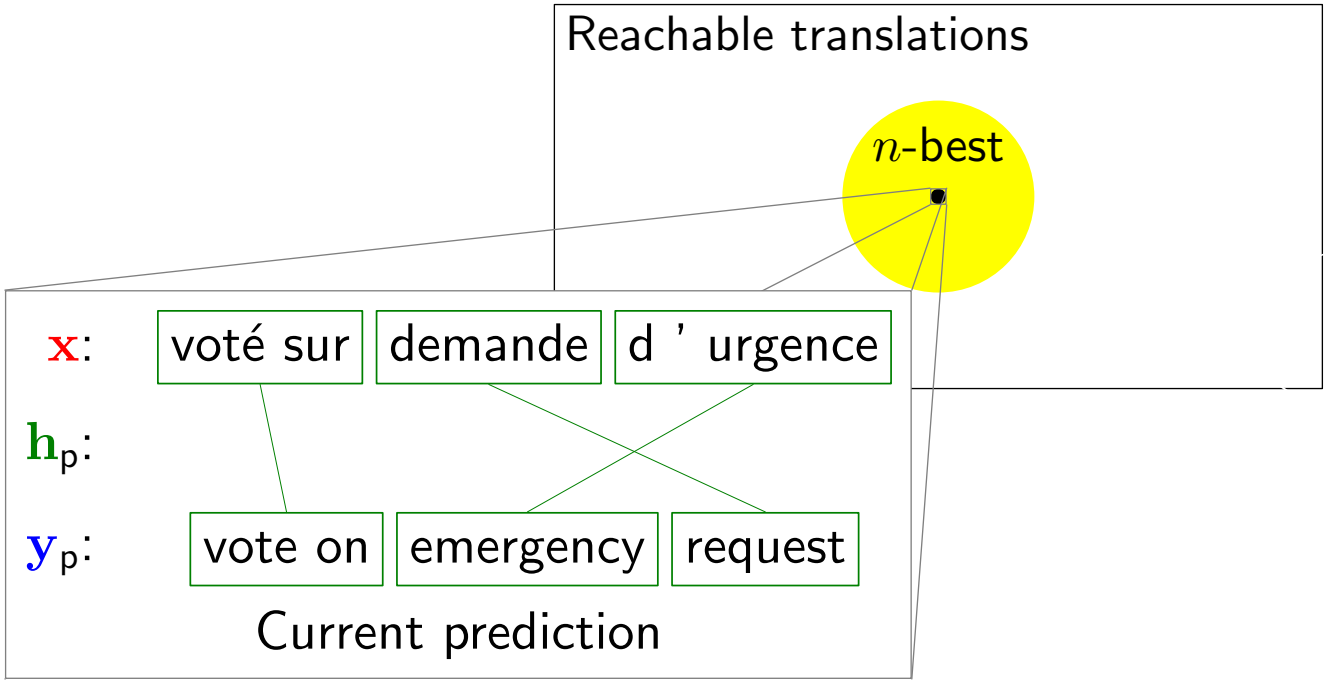
Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure



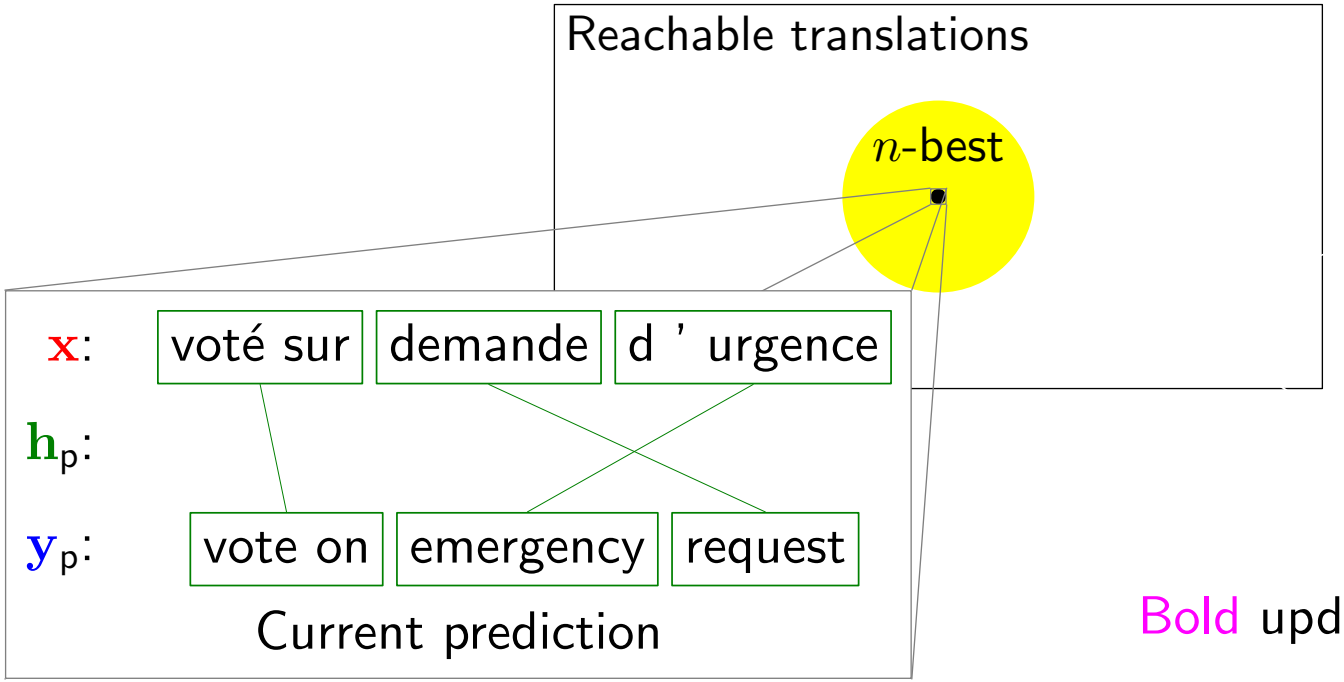
Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure



Bold update: skip example

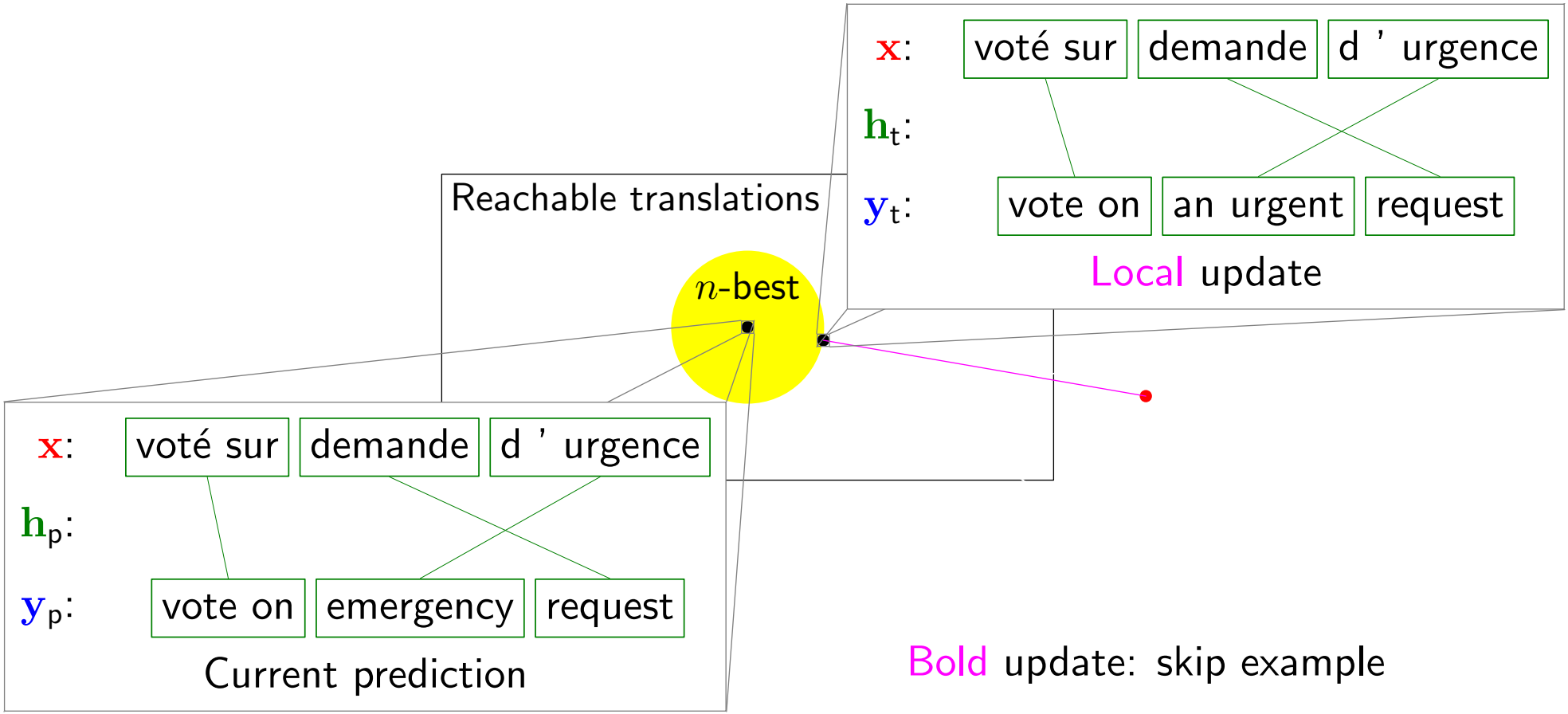
Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure



Bold update: skip example

Update strategies

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(\mathbf{x}, \boxed{\mathbf{y}_t, \mathbf{h}_t}) - \Phi(\mathbf{x}, \mathbf{y}_p, \mathbf{h}_p)$$

Training example (reference)

$\mathbf{x}$: voté sur demande d ' urgence

$\mathbf{y}$: vote on a request for urgent procedure

$\mathbf{x}$:

$\mathbf{h}_t$:

Decoder	Bold	Local
Monotonic	34.3	**34.6**
Limited distortion	33.5	**34.7**

te

$\mathbf{x}$:

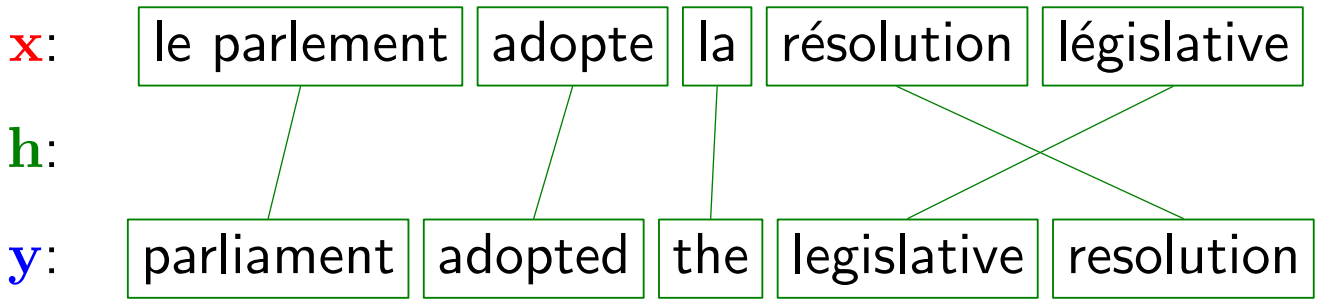
$\mathbf{h}_p$:

$\mathbf{y}_p$:

Current prediction

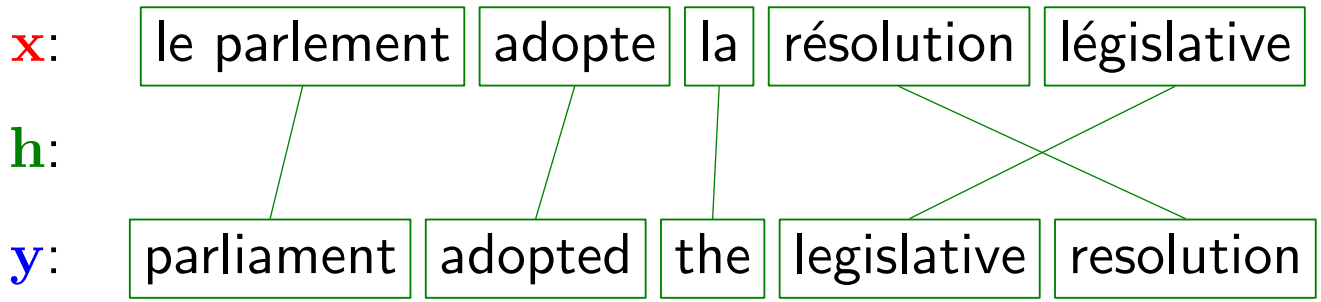
Bold update: skip example

Features



$$\Phi(\mathbf{x}, \mathbf{y}, \mathbf{h}) =$$

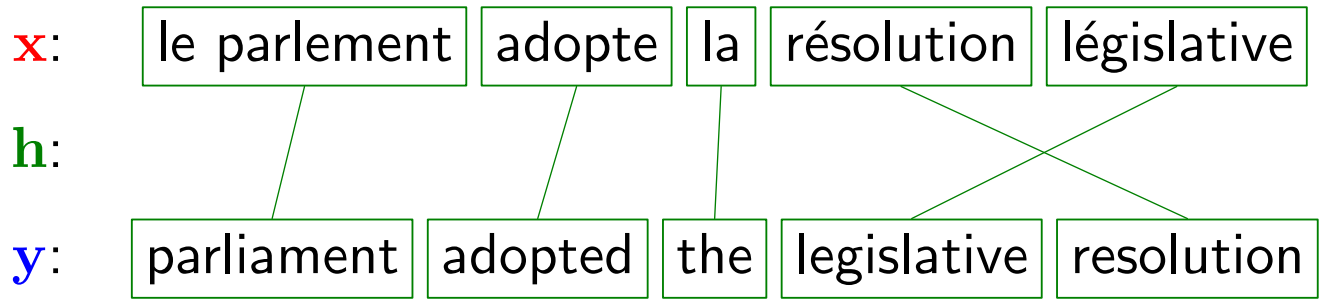
Features



$\Phi(\mathbf{x}, \mathbf{y}, \mathbf{h}) =$

BLANKET	TranslationLogProb	-8.92
	LangModelLogProb	-2.462

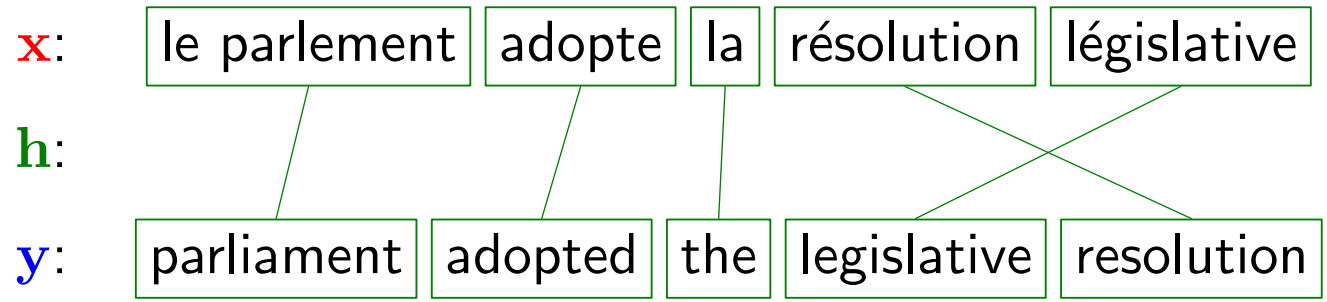
Features



$\Phi(\mathbf{x}, \mathbf{y}, \mathbf{h}) =$

BLANKET	TranslationLogProb	-8.92
	LangModelLogProb	-2.462
LEXICAL	le parlement — parliament	1
	[parliament adopted the]	1
	...	

Features



$\Phi(\mathbf{x}, \mathbf{y}, \mathbf{h}) =$

BLANKET	TranslationLogProb	-8.92
	LangModelLogProb	-2.462
LEXICAL	le parlement — parliament	1
	[parliament adopted the]	1
	...	
POS	DT NN — NN	1
	[NN VBD DT]	1
	...	
	NN JJ / JJ NN (swap)	1

Effect of features

trente-cinq	langues
five	languages

Features	DEV BLEU
BLANKET (untuned)	33.0

Effect of features

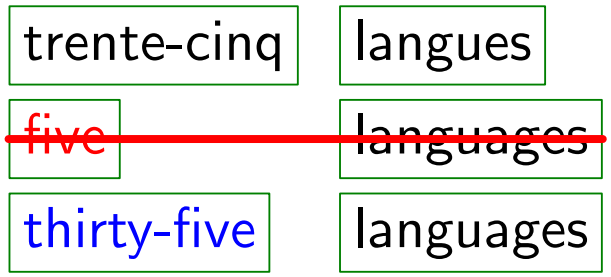
trente-cinq	langues
five	languages

Tune BLANKET features...

- Resulting relative weights: $\mathbf{w}(\text{TranslationLogProb}) = 2$
  $\mathbf{w}(\text{LangModelLogProb}) = 1$

Features	DEV BLEU
BLANKET (untuned)	33.0

Effect of features

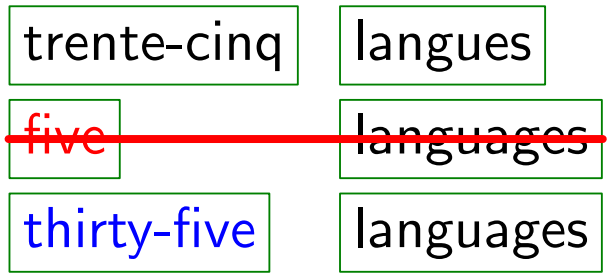


Tune BLANKET features...

- Resulting relative weights: $\mathbf{w}(\text{TranslationLogProb}) = 2$
  $\mathbf{w}(\text{LangModelLogProb}) = 1$

Features	DEV BLEU
BLANKET (untuned)	33.0

Effect of features



Tune BLANKET features...

- Resulting relative weights: $\mathbf{w}(\text{TranslationLogProb}) = 2$
  $\mathbf{w}(\text{LangModelLogProb}) = 1$

Features	DEV BLEU
BLANKET (untuned)	33.0
BLANKET	33.4

Effect of features

pour cela que	j ' ai	voté favorablement	.
for that	i have	voted in favour	.

Features	DEV BLEU
BLANKET (untuned)	33.0
BLANKET	33.4

Effect of features

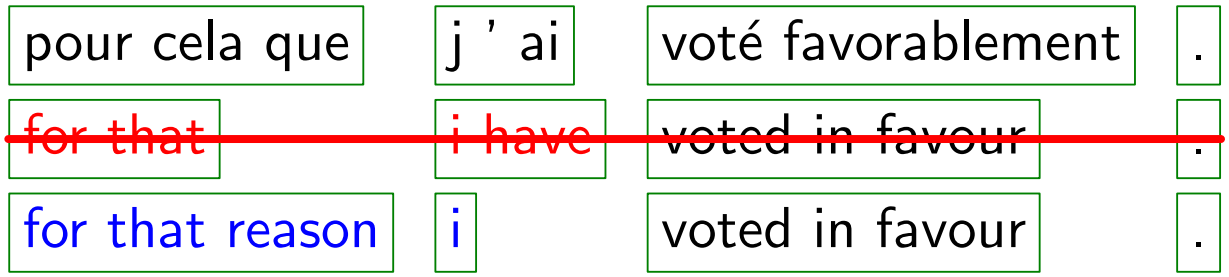
pour cela que	j ' ai	voté favorablement	.
for that	i have	voted in favour	.

Add LEXICAL features...

- j ' ai — i have gets very negative weight
- Literal phrase translations downweighted to allow context-sensitive translations

Features	DEV BLEU
BLANKET (untuned)	33.0
BLANKET	33.4

Effect of features

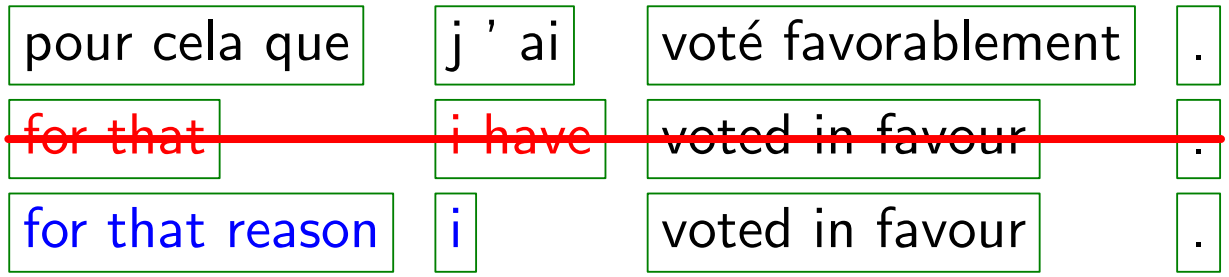


Add LEXICAL features...

- j ' ai — i have gets very negative weight
- Literal phrase translations downweighted to allow context-sensitive translations

Features	DEV BLEU
BLANKET (untuned)	33.0
BLANKET	33.4

Effect of features



Add LEXICAL features...

- j ' ai — i have gets very negative weight
- Literal phrase translations downweighted to allow context-sensitive translations

Features	DEV BLEU
BLANKET (untuned)	33.0
BLANKET	33.4
BLANKET+LEXICAL	35.0

Effect of features

How can we generalize beyond lexical features?

Features	Dev BLEU
Blanket (untuned)	33.0
Blanket	33.4
Blanket+Lexical	35.0

Effect of features

How can we generalize beyond lexical features?

Add POS features...

- la réalisation du droit — the right has negative weight

  generalizes to DT NN IN NN — DT NN
- Number of nonzero features drops (1.55M to 1.24M)

Features	Dev BLEU
Blanket (untuned)	33.0
Blanket	33.4
Blanket+Lexical	35.0

Effect of features

How can we generalize beyond lexical features?

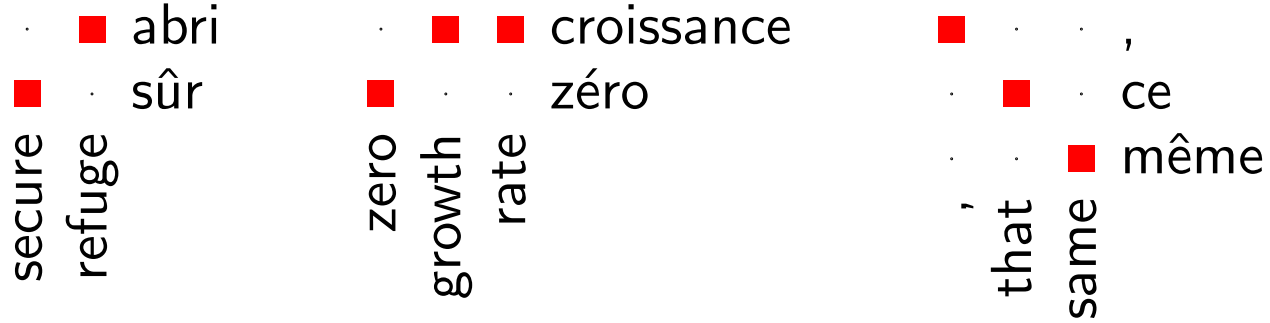
Add POS features...

- la réalisation du droit — the right has negative weight generalizes to DT NN IN NN — DT NN
- Number of nonzero features drops (1.55M to 1.24M)

Features	Dev BLEU
Blanket (untuned)	33.0
Blanket	33.4
Blanket+Lexical	35.0
Blanket+Lexical+POS	35.3

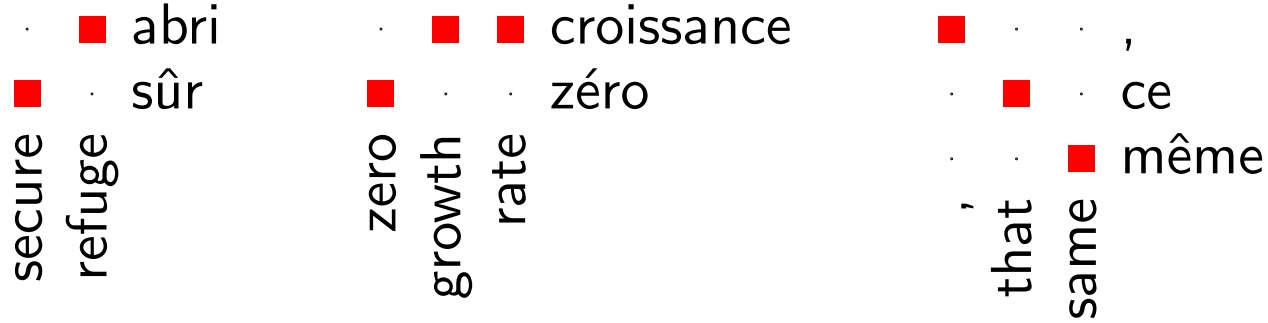
Alignment constellation features

Phrase-extraction heuristic is important [Koehn '03]



Alignment constellation features

Phrase-extraction heuristic is important [Koehn '03]

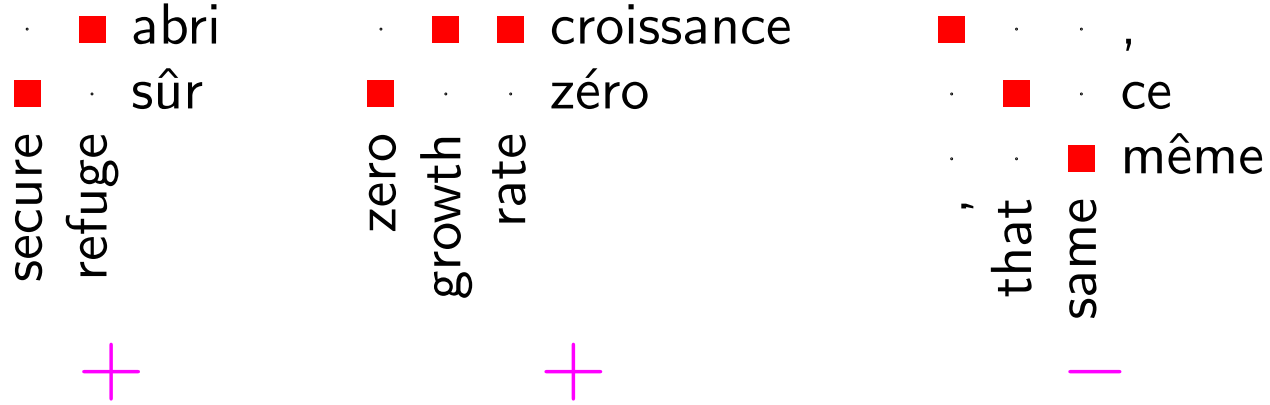


Put features on the phrase-extraction process itself

$$\Phi(\mathbf{x}, \mathbf{y}, \mathbf{h}) = \begin{cases} \boxed{\blacksquare \cdot \\ \cdot \blacksquare} & 1 \\ \ldots & \end{cases}$$

Alignment constellation features

Phrase-extraction heuristic is important [Koehn '03]

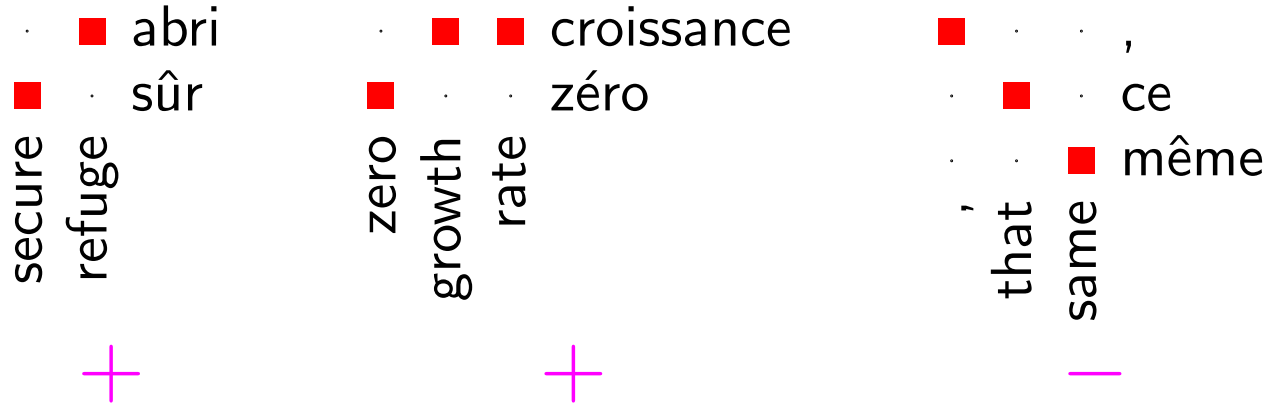


Put features on the phrase-extraction process itself

$$\Phi(\mathbf{x}, \mathbf{y}, \mathbf{h}) = \begin{cases} \boxed{\blacksquare \cdot \\ \cdot \blacksquare} & 1 \\ \ldots & \end{cases}$$

Alignment constellation features

Phrase-extraction heuristic is important [Koehn '03]



Put features on the phrase-extraction process itself

$$\Phi(\mathbf{x}, \mathbf{y}, \mathbf{h}) = \begin{cases} \boxed{\blacksquare \cdot \\ \cdot \blacksquare} & 1 \\ \ldots & \end{cases}$$

Features	-Const	+Const
Blanket	31.8	32.2
Blanket+Lexical	32.2	32.5
Blanket+Lexical+POS	32.3	32.5

Summary of results

Monotonic systems	Test BLEU
Pharaoh (MERT)	28.8
Blanket	28.4
Blanket+Lexical+POS	**29.6**

Summary of results

Monotonic systems	Test BLEU
Pharaoh (MERT)	28.8
Blanket	28.4
Blanket+Lexical+POS	**29.6**

Distortion systems	Test BLEU
Pharaoh (MERT) [Full]	29.5
Blanket [Limited]	30.0
Blanket+Lexical+POS [Limited]	**30.9**

Conclusions

- Perceptron training with hidden variables: local updates are better

Conclusions

- Perceptron training with hidden variables:
  local updates are better

- Allow many expressive features:
  blanket, lexical, POS, alignment constellation

Conclusions

- Perceptron training with hidden variables:
  local updates are better

- Allow many expressive features:
  blanket, lexical, POS, alignment constellation

- Significant BLEU improvements over Pharaoh

Conclusions

- Perceptron training with hidden variables:
  local updates are better

- Allow many expressive features:
  blanket, lexical, POS, alignment constellation

- Significant BLEU improvements over Pharaoh

- Extension: beyond phrase-based models

Conclusions

- Perceptron training with hidden variables:
  local updates are better

- Allow many expressive features:
  blanket, lexical, POS, alignment constellation

- Significant BLEU improvements over Pharaoh

- Extension: beyond phrase-based models

That's it.