

Tailoring Word Alignments to Syntactic Machine Translation

John DeNero and Dan Klein



Presentation and paper:

<http://nlp.cs.berkeley.edu/pages/WordAligner.html>



Tailoring Word Alignments to Syntactic Machine Translation

- Setting:** Syntactic MT with tree transducers
- Problem:** Alignment errors that contradict constituent structure impede the rule extraction process
- Proposal:** Condition word alignment on syntactic structure



Translating with Tree Transducers

Source: Les emplois sont axés sur la carrière .



Translating with Tree Transducers

Source: Les emplois sont axés sur la carrière .

Gloss: *The jobs are centered on the career .*

Translating with Tree Transducers

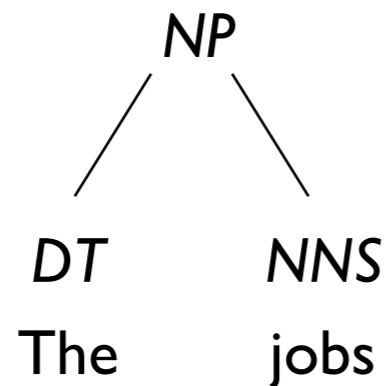
Source: Les emplois sont axés sur la carrière .

Gloss: *The jobs are centered on the career .*

Transducer rule: $(NP (DT The) (NNS jobs)) \Rightarrow$ Les emplois

Translating with Tree Transducers

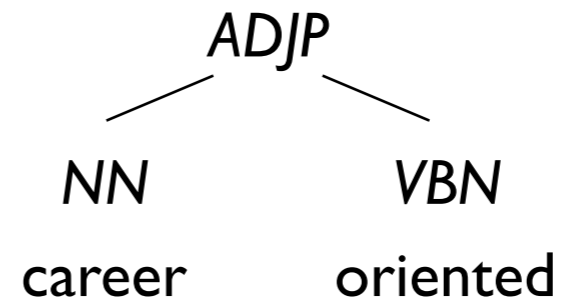
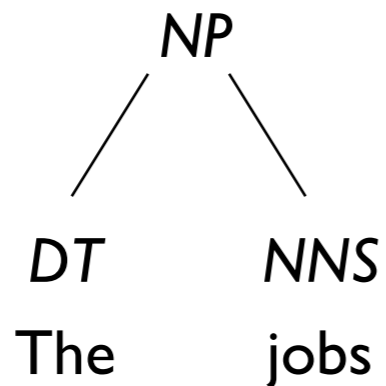
	<i>NP</i>					
Source:	Les	emplois	sont	axés	sur	la carrière .
Gloss:	<i>The</i>	<i>jobs</i>	<i>are</i>	<i>centered</i>	<i>on</i>	<i>the career .</i>



Transducer rule: $(NP (DT The) (NNS jobs)) \Rightarrow$ Les emplois

Translating with Tree Transducers

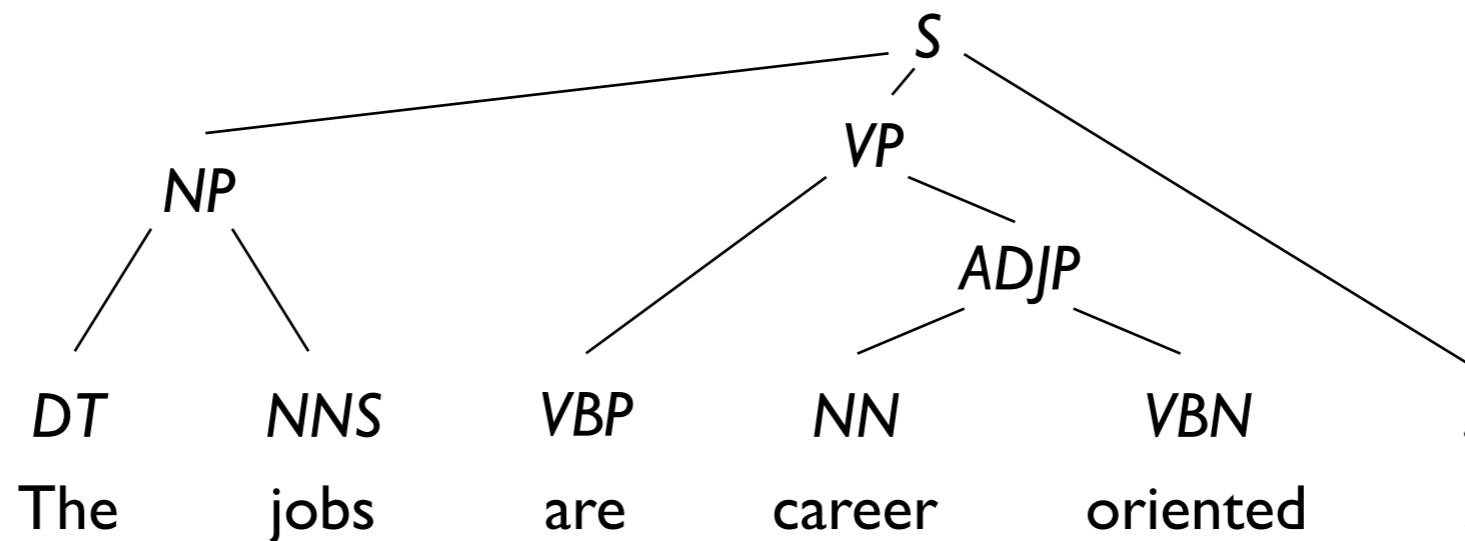
	<i>NP</i>		<i>ADJP</i>	
Source:	Les emplois	sont	axés sur la carrière	.
Gloss:	The jobs	are	centered on the career	.



Transducer rule: (*ADJP* (*NN* career) (*VBN* oriented)) => axés sure la carrière

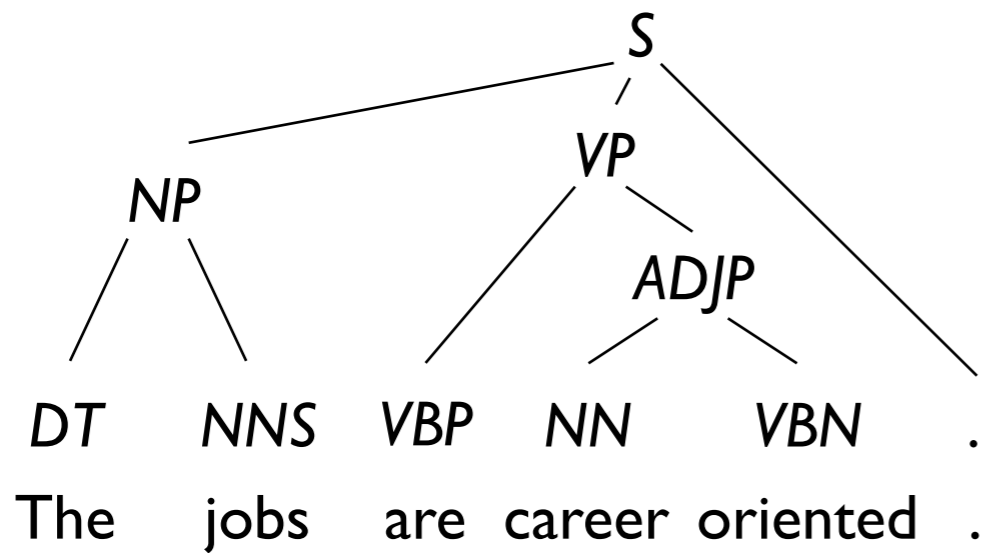
Translating with Tree Transducers

	<i>NP</i>			<i>ADJP</i>				
Source:	Les	emplois	sont	axés	sur	la	carrière	.
Gloss:	The	jobs	are	centered	on	the	career	.



Transducer rule: $(S \ NP_1 \ (VP \ (VBP \ are) \ ADJP_2) \ (.\ .)) \Rightarrow \ NP_1 \ sont \ ADJP_2 \ .$

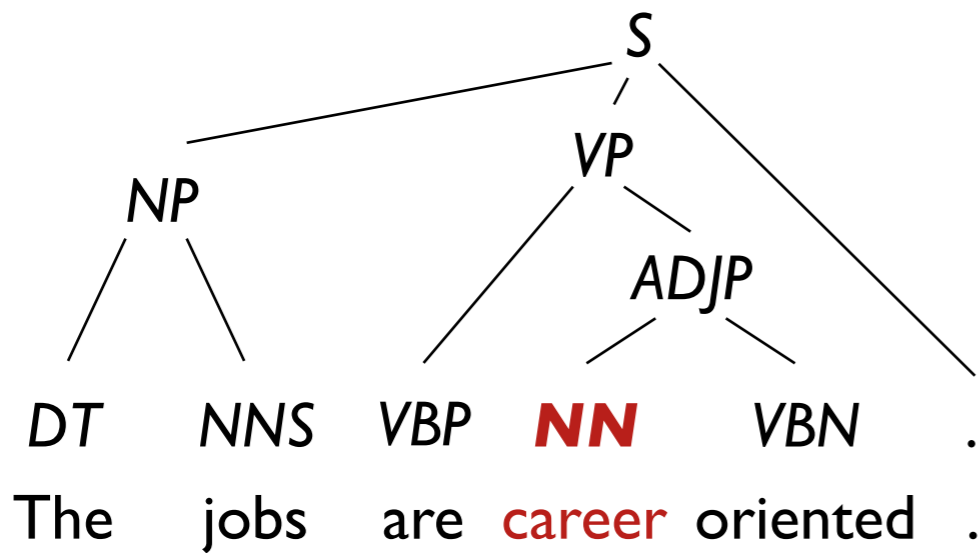
Extracting Transducer Rules



Extraction Procedure (Galley et al., '04 & '06)

						Les
						emplois
						sont
						axés
						sur
						la
						carrière
						.

Extracting Transducer Rules

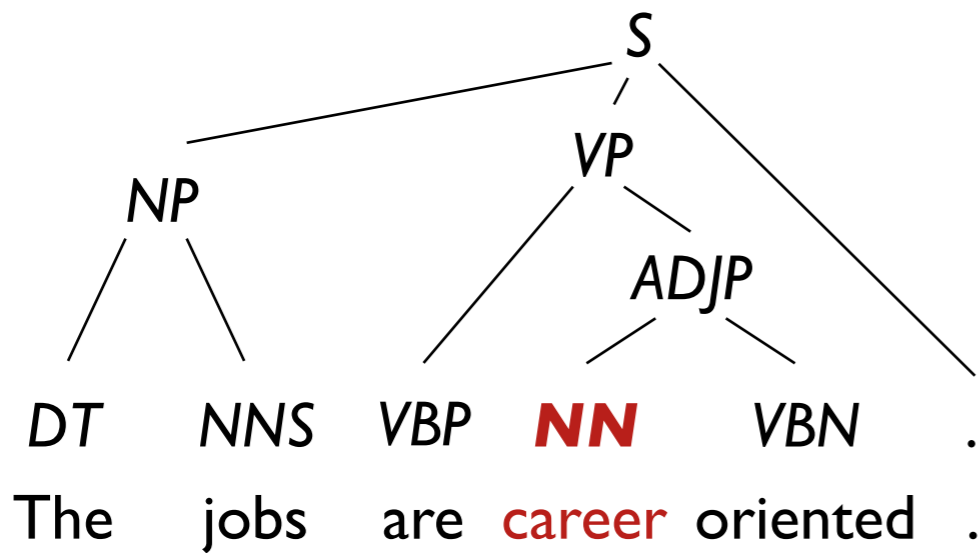


Extraction Procedure (Galley et al., '04 & '06)

I. Choose a constituent

						Les
						emplois
						sont
						axés
						sur
						la
						carrière
						.

Extracting Transducer Rules

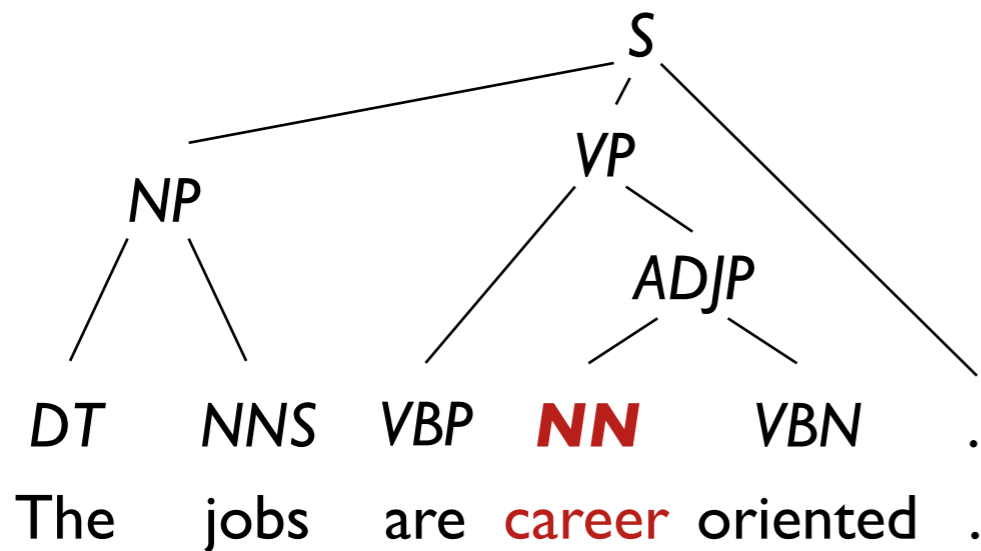


Les
emplois
sont
axés
sur
la
carrière
.

Extraction Procedure (Galley et al., '04 & '06)

1. Choose a constituent
2. Choose a region around constituent alignments

Extracting Transducer Rules

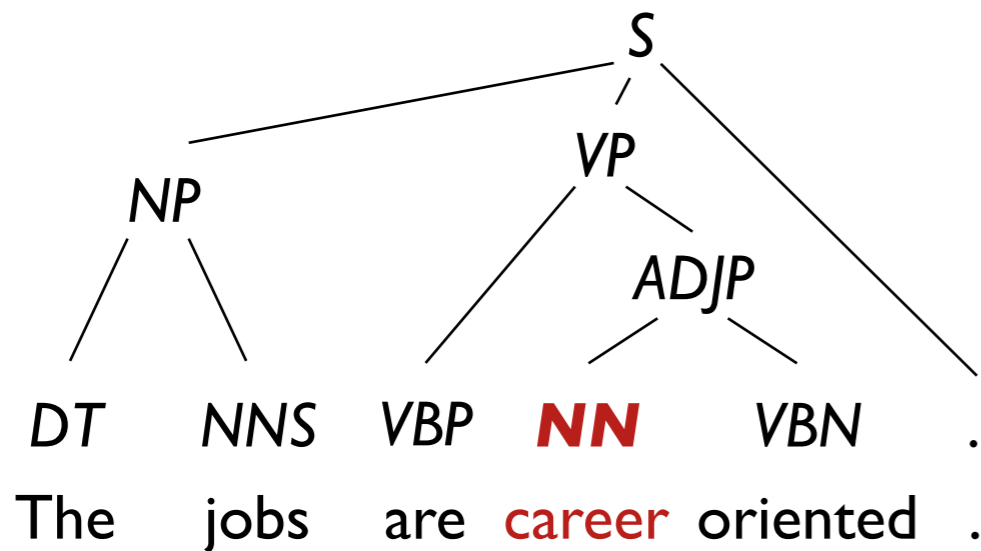


Les
 emplois
 sont
 axés
 sur
 la
 carrière
 .

Extraction Procedure (Galley et al., '04 & '06)

1. Choose a constituent
2. Choose a region around constituent alignments
3. Verify that alignment is consistent with region

Extracting Transducer Rules

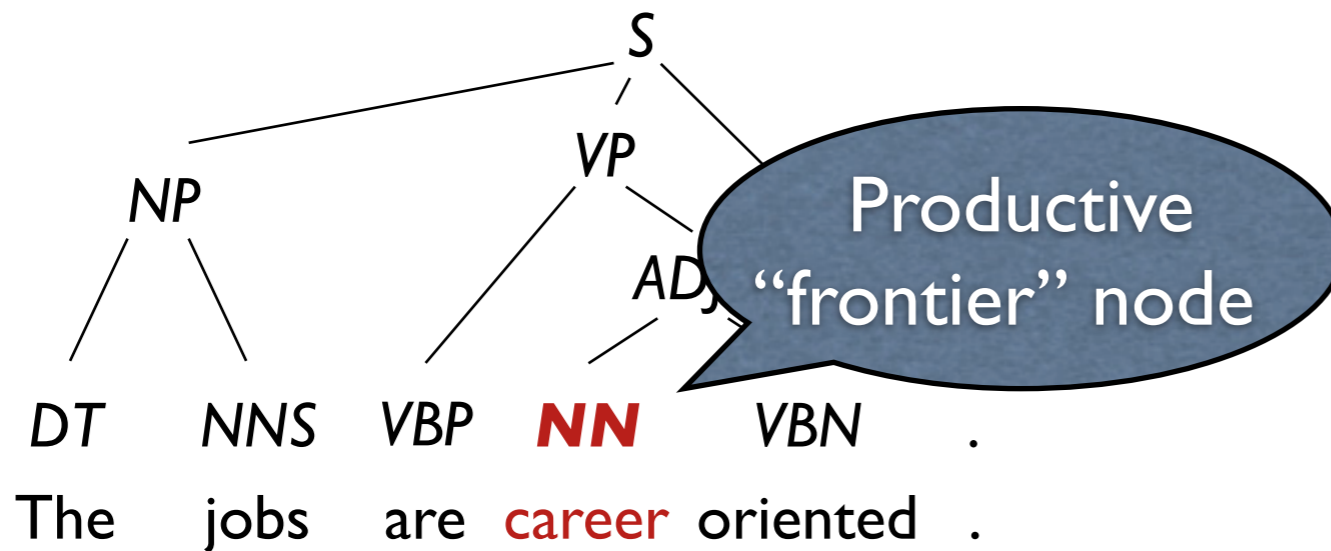


Les
emplois
sont
axés
sur
la
carrière
.

Extraction Procedure (Galley et al., '04 & '06)

1. Choose a constituent
2. Choose a region around constituent alignments
3. Verify that alignment is consistent with region
4. Extract phrase:
(**NN career**) => **carrière**

Extracting Transducer Rules

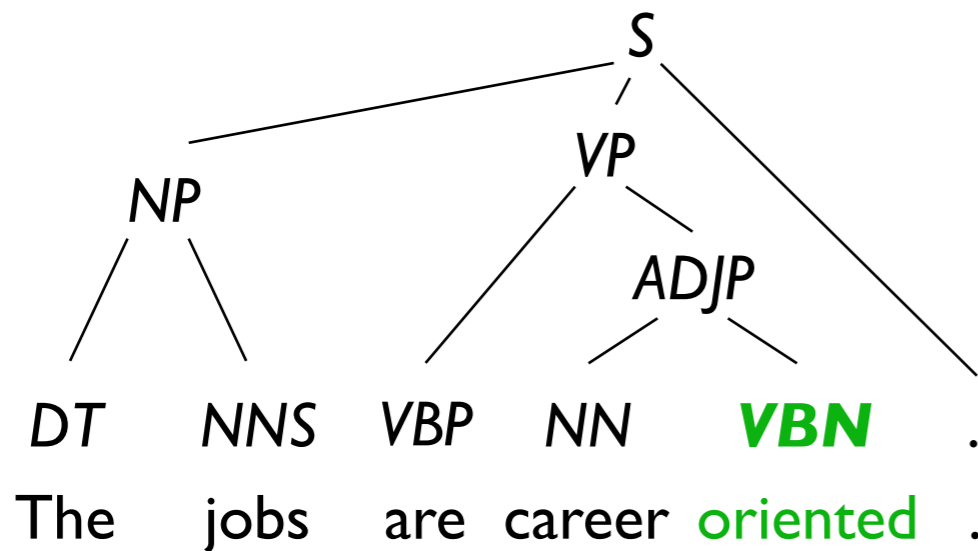


Les
emplois
sont
axés
sur
la
carrière
.

Extraction Procedure (Galley et al., '04 & '06)

1. Choose a constituent
2. Choose a region around constituent alignments
3. Verify that alignment is consistent with region
4. Extract phrase:
(**NN career**) => **carrière**

Extracting Transducer Rules

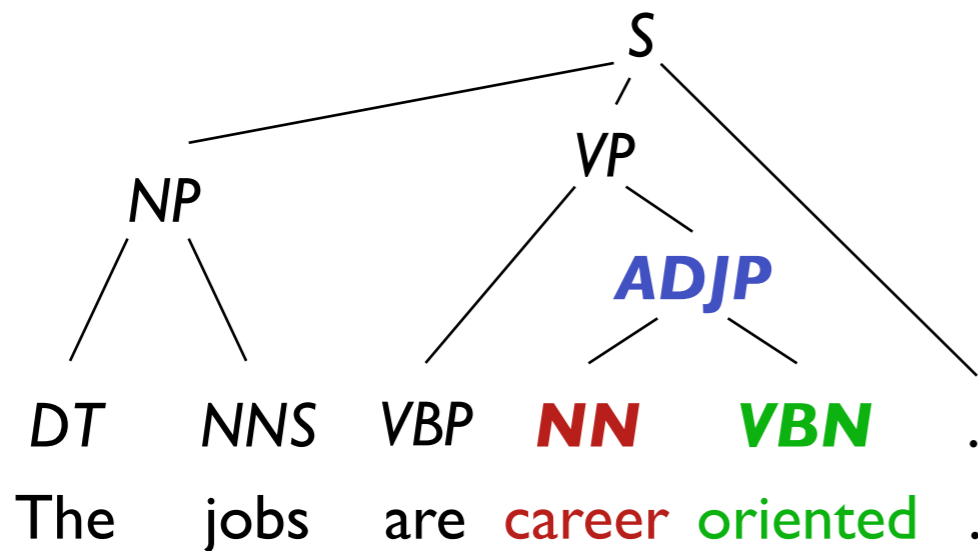


Les
emplois
sont
axés
sur
la
carrière
.

Extraction Procedure (Galley et al., '04 & '06)

1. Choose a constituent
2. Choose a region around constituent alignments
3. Verify that alignment is consistent with region
4. Extract phrase:
(*VBN oriented*) => *axés sur*

Extracting Transducer Rules

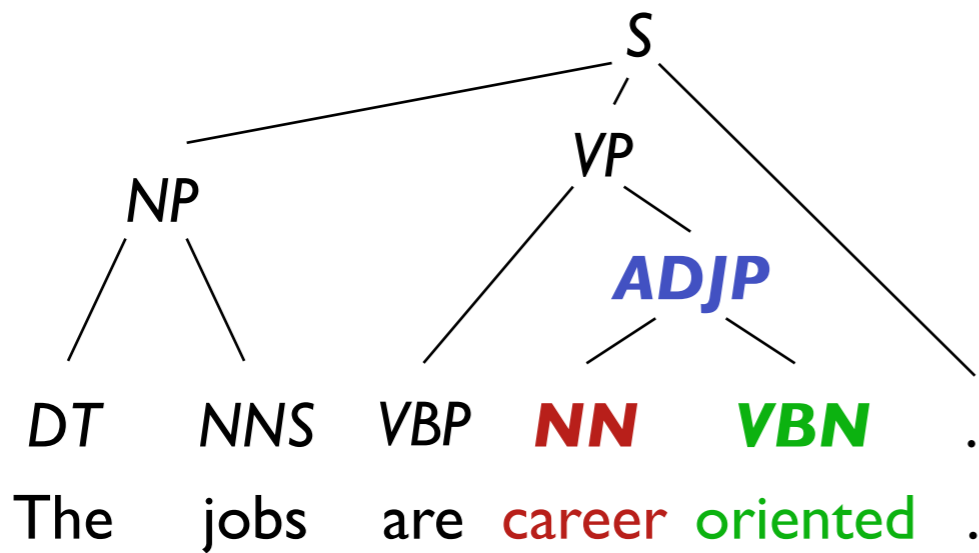


Les
emplois
sont
axés
sur
la
carrière
.

Extraction Procedure (Galley et al., '04 & '06)

1. Choose a constituent
2. Choose a region around constituent alignments
3. Verify that alignment is consistent with region
4. Extract phrase:
(ADJP NN₁ VBN₂) ⇒ VBN₂ la NN₁

Extracting Transducer Rules

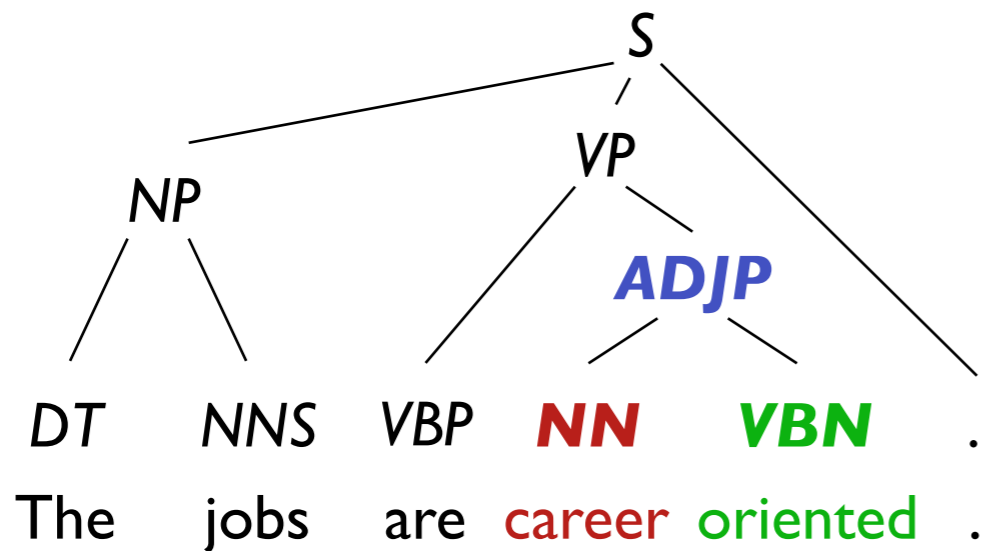


Rule Combination (Galley et al., '06)

(ADJP NN₁ VBN₂) => VBN₂ la NN₁
 (VBN oriented) => axés sur
 (NN career) => carrière

Les
emplois
sont
axés
sur
la
carrière
.

Extracting Transducer Rules



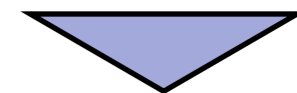
Les
 emplois
 sont
 axés
 sur
 la
 carrière
 .

Rule Combination (Galley et al., '06)

$(ADJP\ NN_1\ VBN_2) \Rightarrow VBN_2\ la\ NN_1$

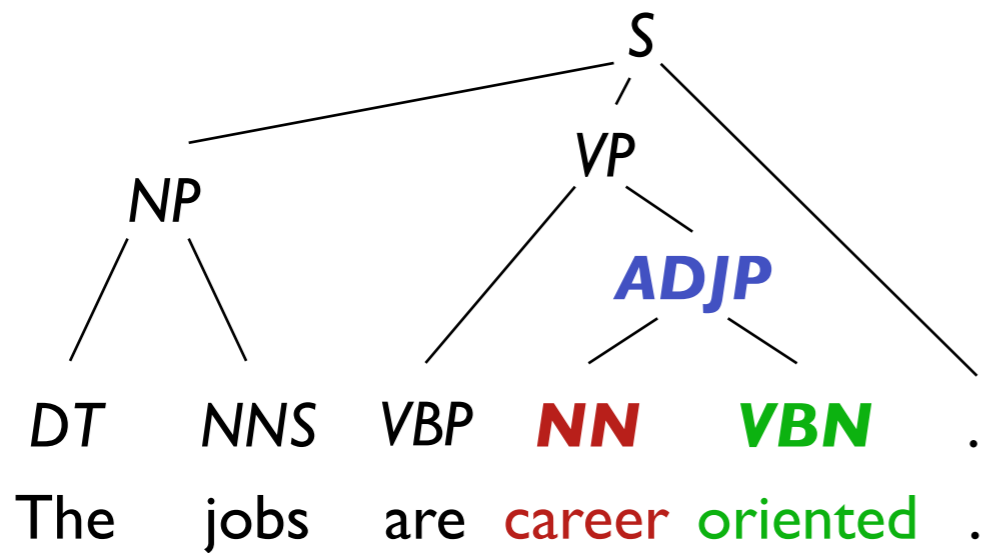
$(VBN\ oriented) \Rightarrow axés\ sur$

$(NN\ career) \Rightarrow carrière$



$(ADJP\ NN_1\ (VBN\ oriented))$
 $\Rightarrow axés\ sur\ la\ NN_1$

Extracting Transducer Rules



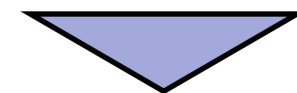
Les
 emplois
 sont
 axés
 sur
 la
 carrière
 .

Rule Combination (Galley et al., '06)

$(ADJP\ NN_1\ VBN_2) \Rightarrow VBN_2\ la\ NN_1$

$(VBN\ oriented) \Rightarrow axés\ sur$

$(NN\ career) \Rightarrow carrière$



$(ADJP\ NN_1\ (VBN\ oriented))$

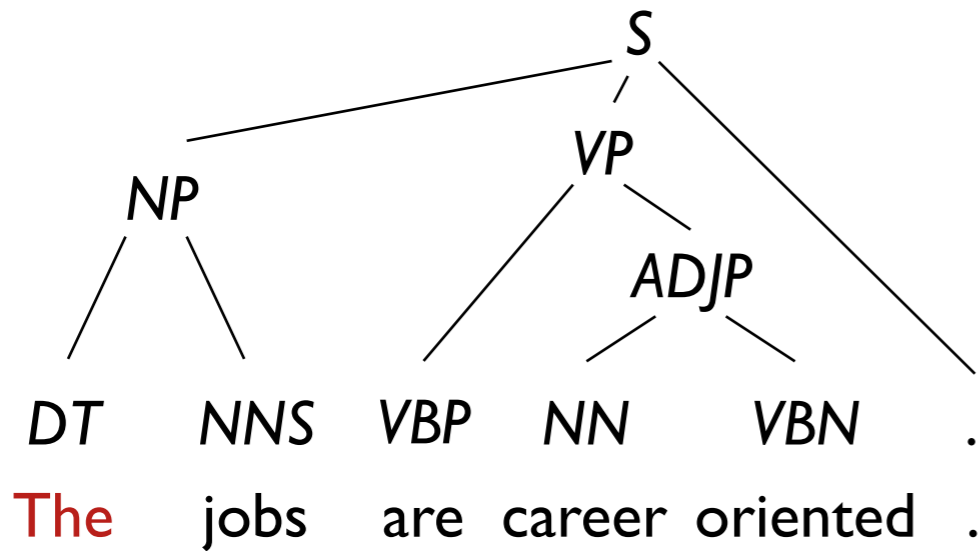
$\Rightarrow axés\ sur\ la\ NN_1$

$(ADJP\ (NN\ career)\ (VBN\ oriented))$

$\Rightarrow axés\ sur\ la\ carrière$

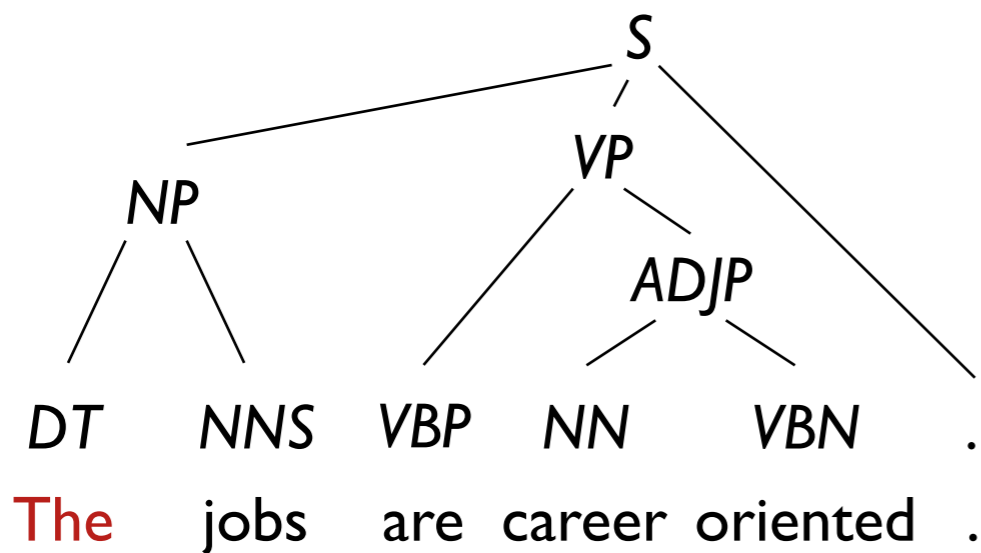
...

Alignment Errors that Prevent Rule Extraction



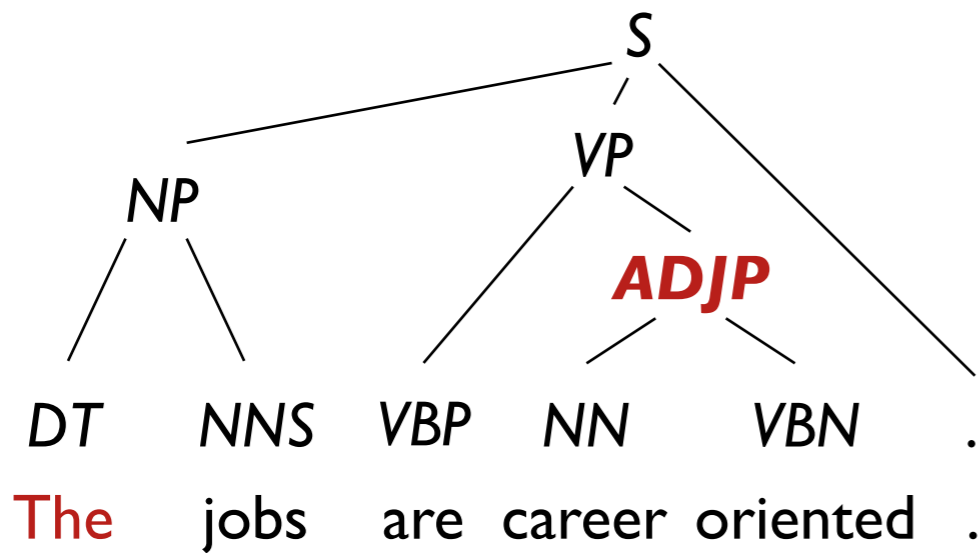
						Les
						emplois
						sont
						axés
						sur
						<i>la</i>
						carrière
						.

Alignment Errors that Prevent Rule Extraction



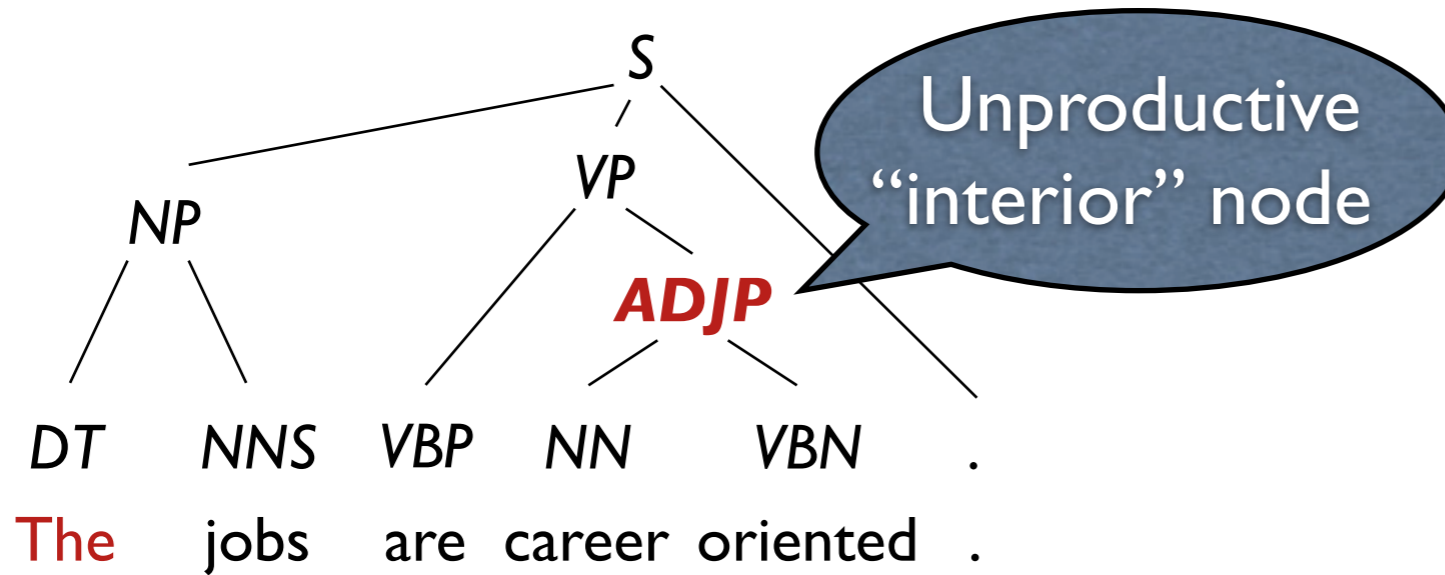
						Les
						emplois
						sont
						axés
						sur
						la
						carrière
						.

Alignment Errors that Prevent Rule Extraction



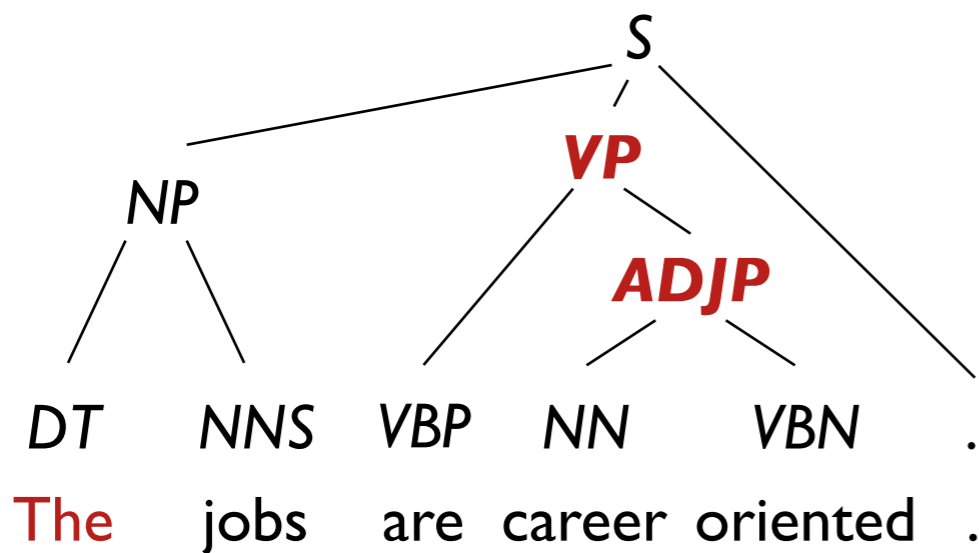
						Les
						emplois
						sont
						axés
						sur
						la
						carrière
						.

Alignment Errors that Prevent Rule Extraction



						Les
						emplois
						sont
						axés
						sur
						la
						carrière
						.

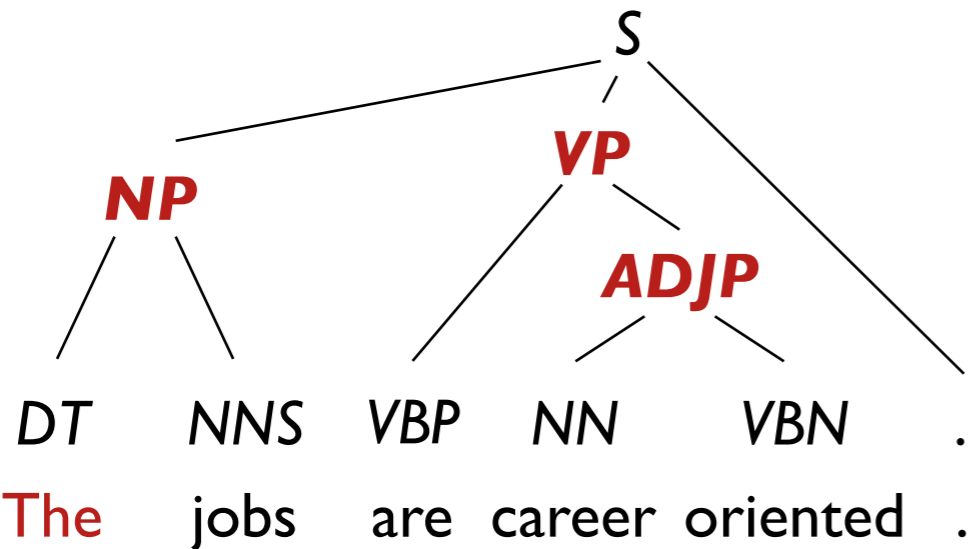
Alignment Errors that Prevent Rule Extraction



						Les
						emplois
						sont
						axés
						sur
						la
						carrière
						.

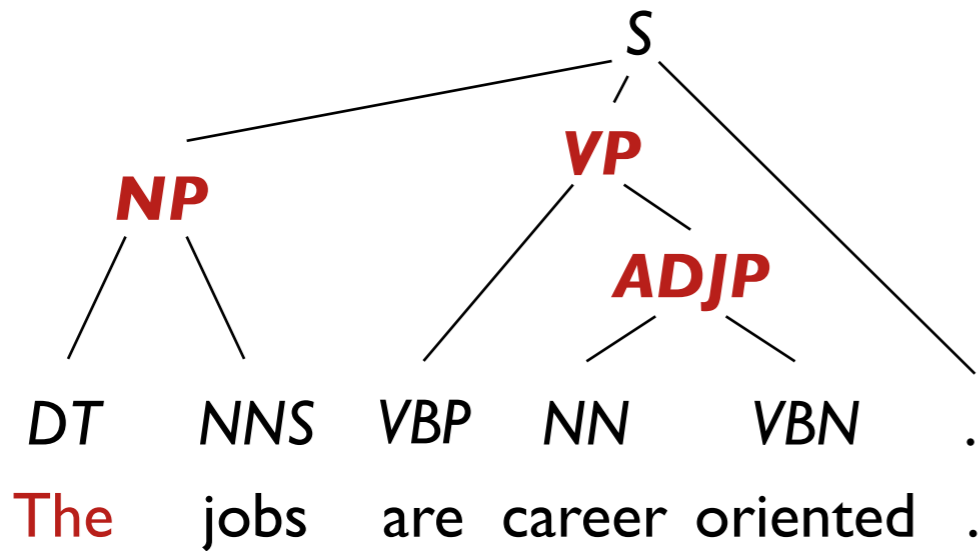
Detailed description: A 8x6 grid representing word alignment between English and French. The English words are in the top row: 'The', 'jobs', 'are', 'career', 'oriented', '.'. The French words are in the right column: 'Les', 'emplois', 'sont', 'axés', 'sur', 'la', 'carrière', '.'. Blue dashed lines indicate alignment: a vertical line from 'The' to 'la', a vertical line from 'jobs' to 'emplois', a vertical line from 'are' to 'sont', a vertical line from 'career' to 'carrière', and a vertical line from 'oriented' to 'axés'. A horizontal dashed line is drawn between the 'are' and 'career' columns. A red shaded cell is located at the intersection of the 'The' column and the 'la' row.

Alignment Errors that Prevent Rule Extraction



						Les
						emplois
						sont
						axés
						sur
						la
						carrière
						.

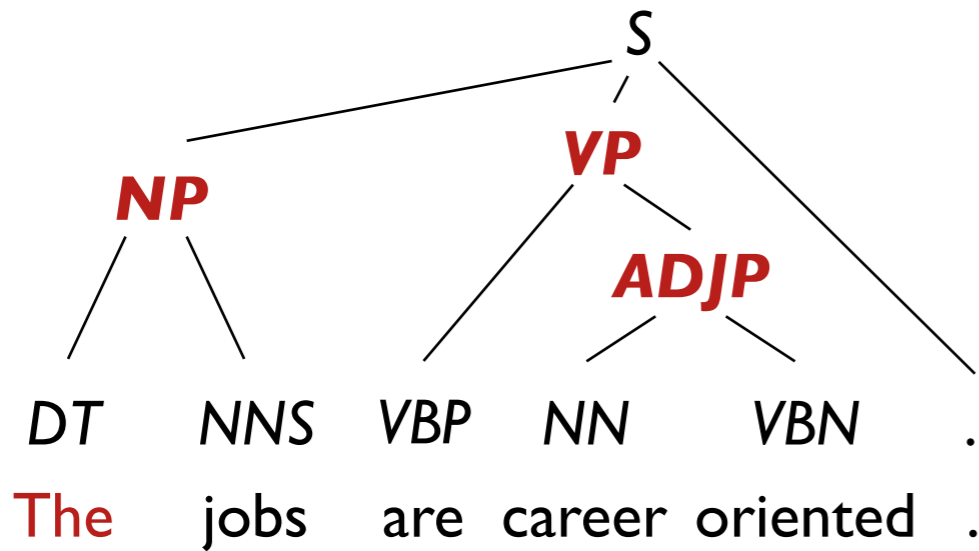
Alignment Errors that Prevent Rule Extraction



Net effect on extraction:

						Les
						emplois
						sont
						axés
						sur
						la
						carrière
						.

Alignment Errors that Prevent Rule Extraction

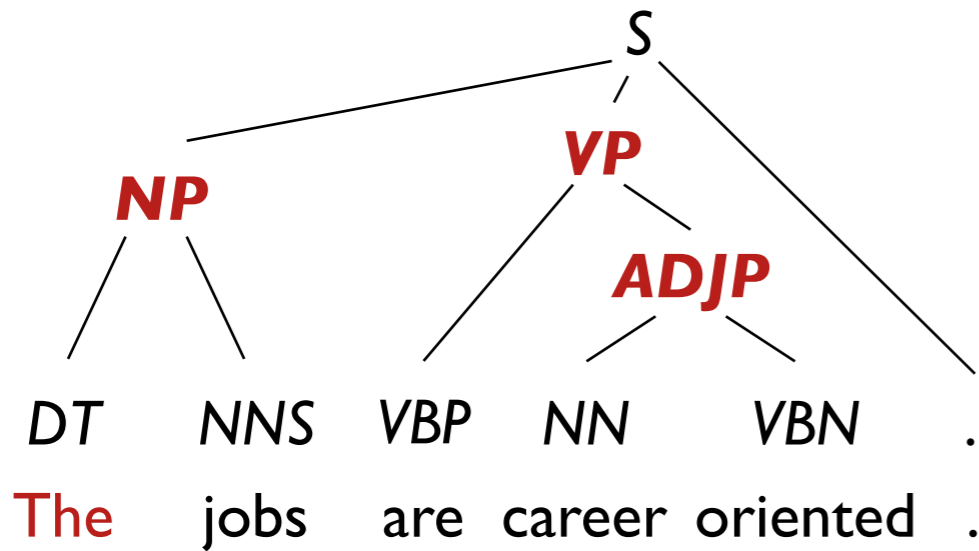


Net effect on extraction:

- 2 instead of 7 recursive rules can be extracted

						Les
						emplois
						sont
						axés
						sur
						la
						carrière
						.

Alignment Errors that Prevent Rule Extraction

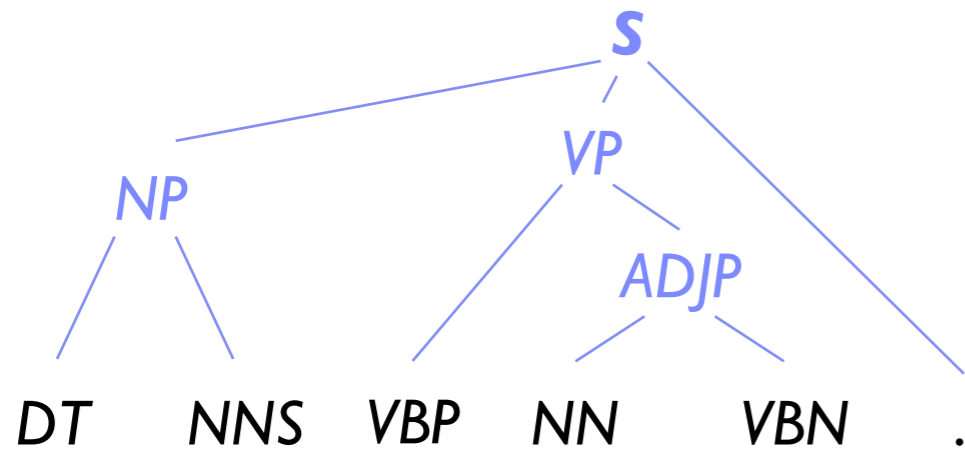


Les
emplois
sont
axés
sur
la
carrière
.

Net effect on extraction:

- 2 instead of 7 recursive rules can be extracted
- Smallest recursive rule that *can* be extracted:

Alignment Errors that Prevent Rule Extraction



The jobs are career oriented .

						Les
						emplois
						sont
						axés
						sur
						la
						carrière
						.

Net effect on extraction:

- 2 instead of 7 recursive rules can be extracted
- Smallest recursive rule that can be extracted:

(S (NP (DT The) NNS₂)
 (VP VBP₃
 (ADJP NN₄ VBN₅))
 .6)
 => Les NNS₂ VBP₃ VBN₅ NN₄ .6

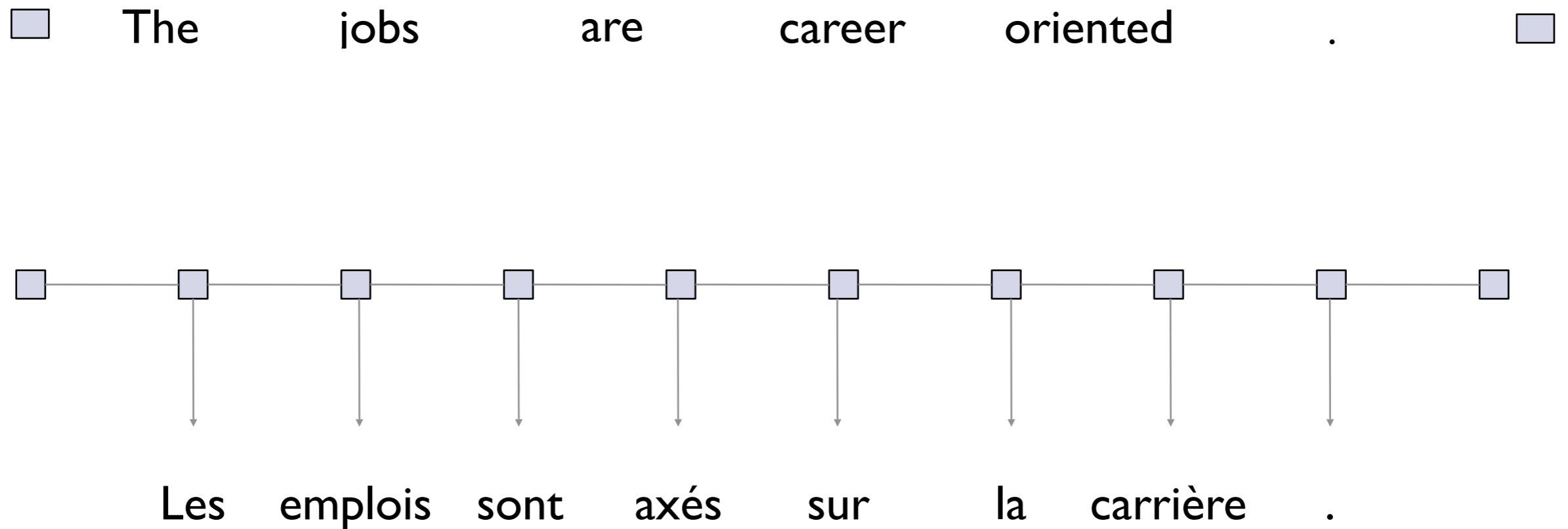


Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j|e_{a_j}) \cdot p(a_j|a_{j-})$$

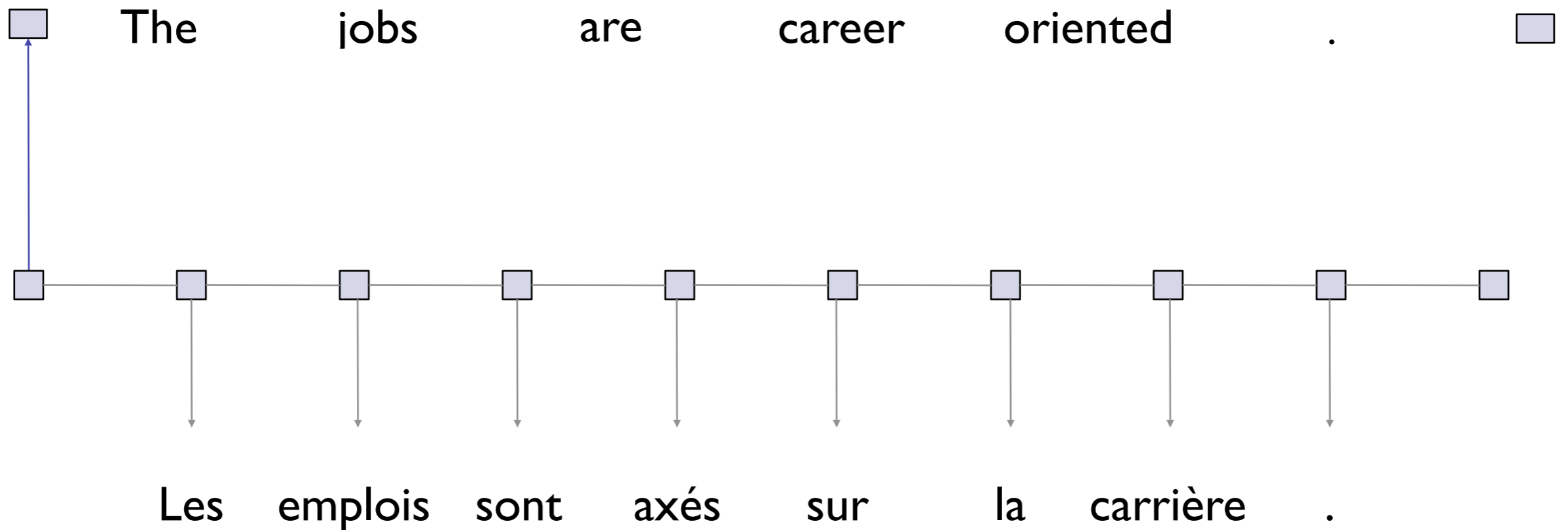
Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$



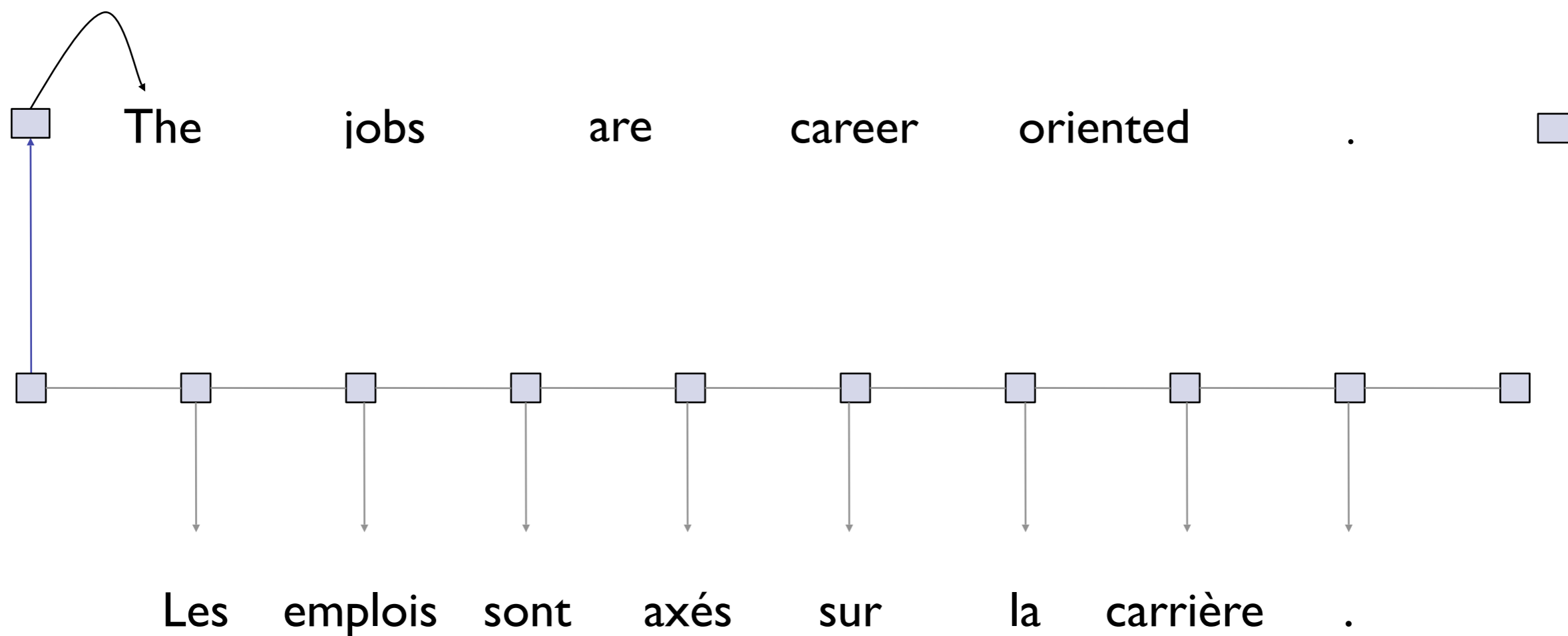
Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$



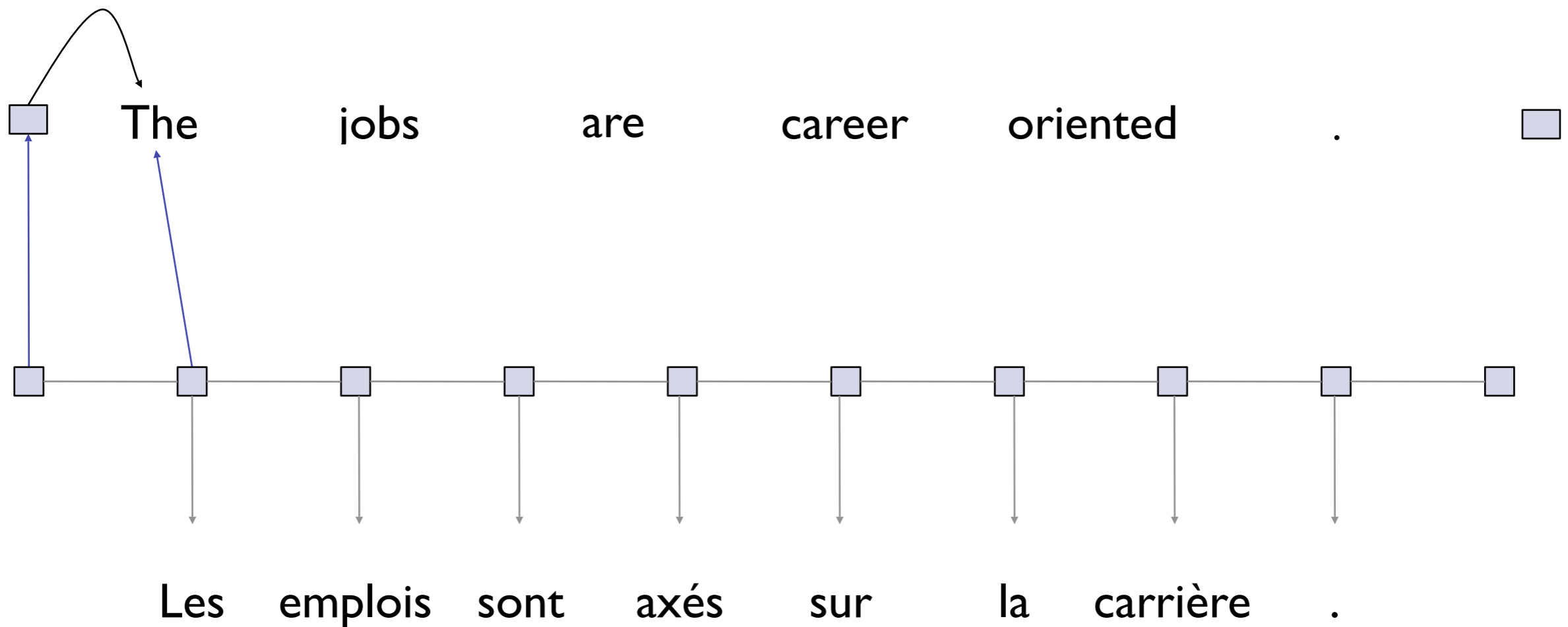
Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$



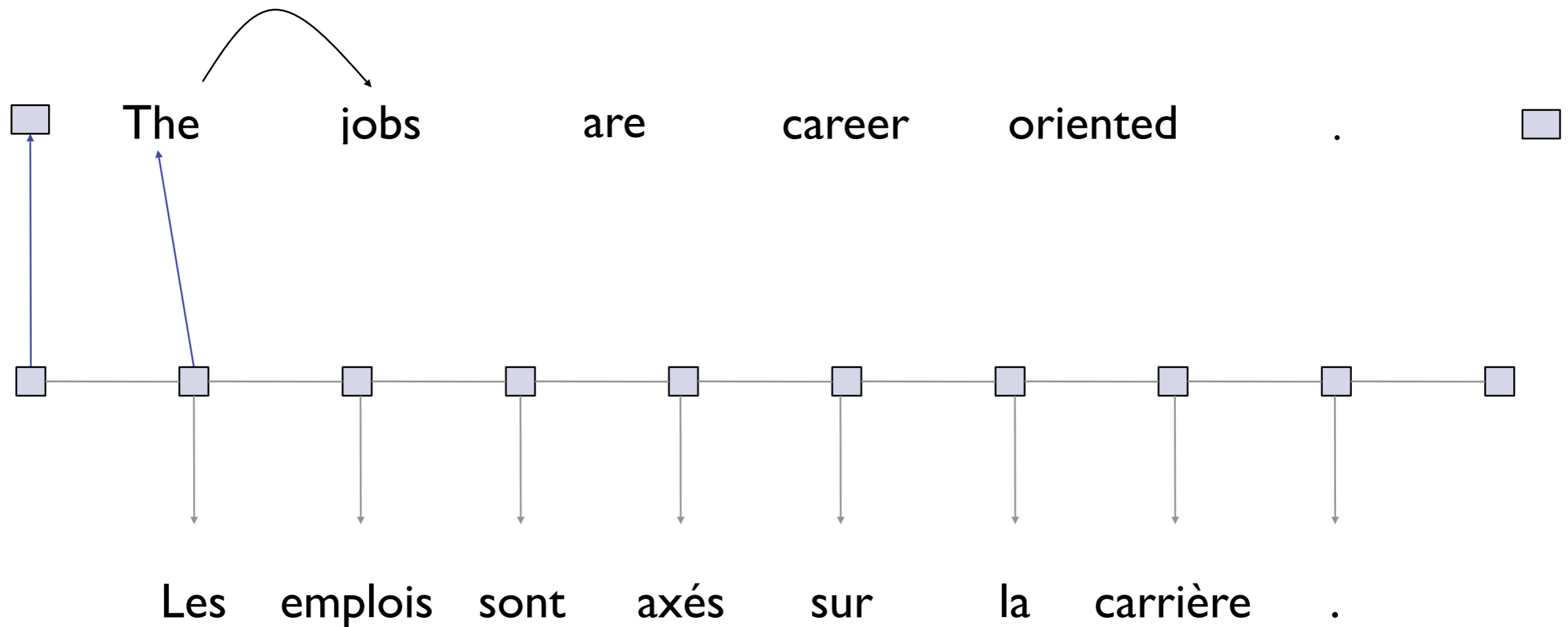
Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$



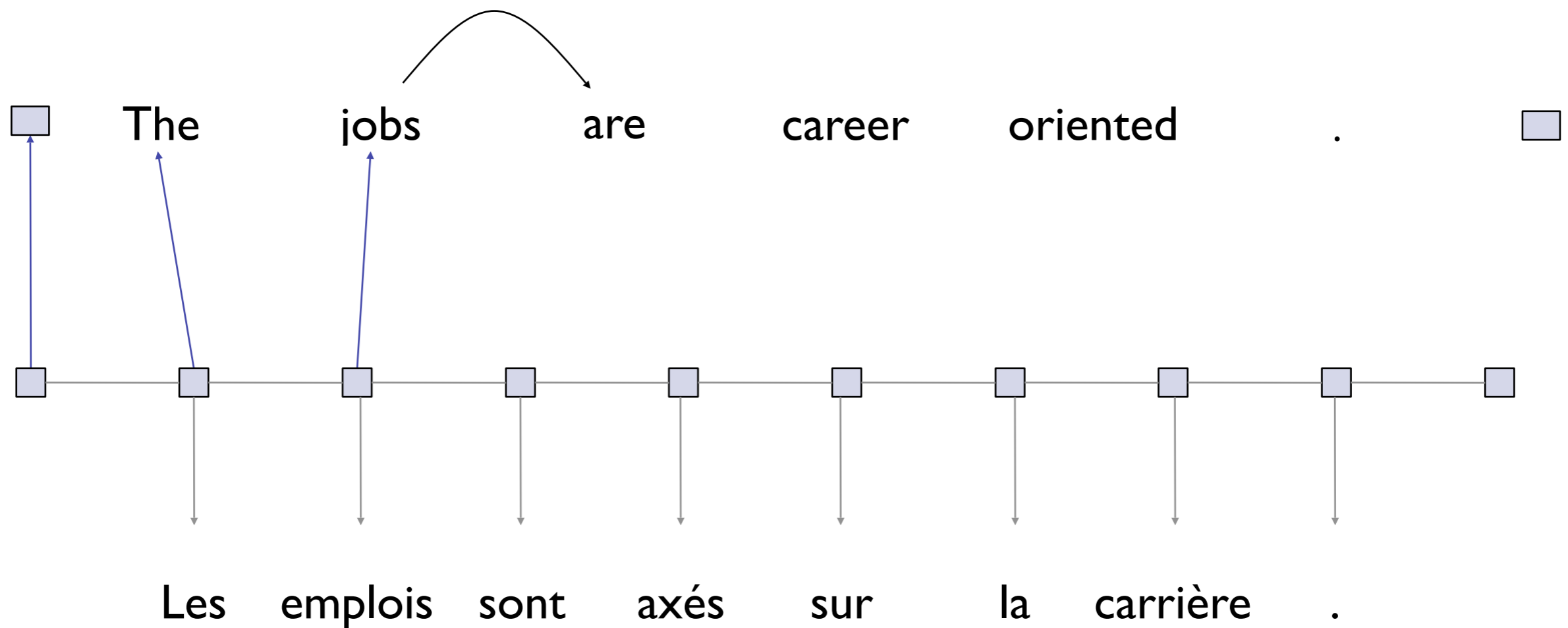
Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$



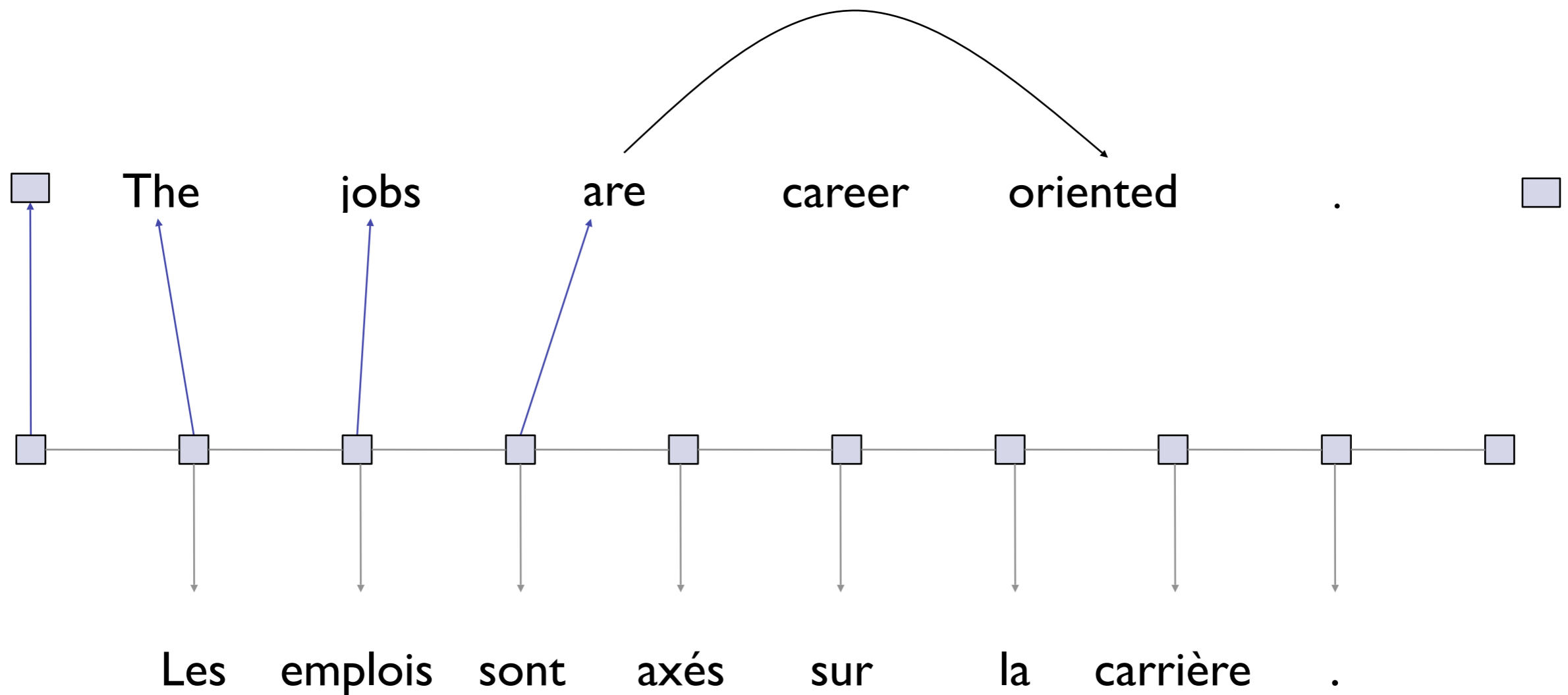
Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$



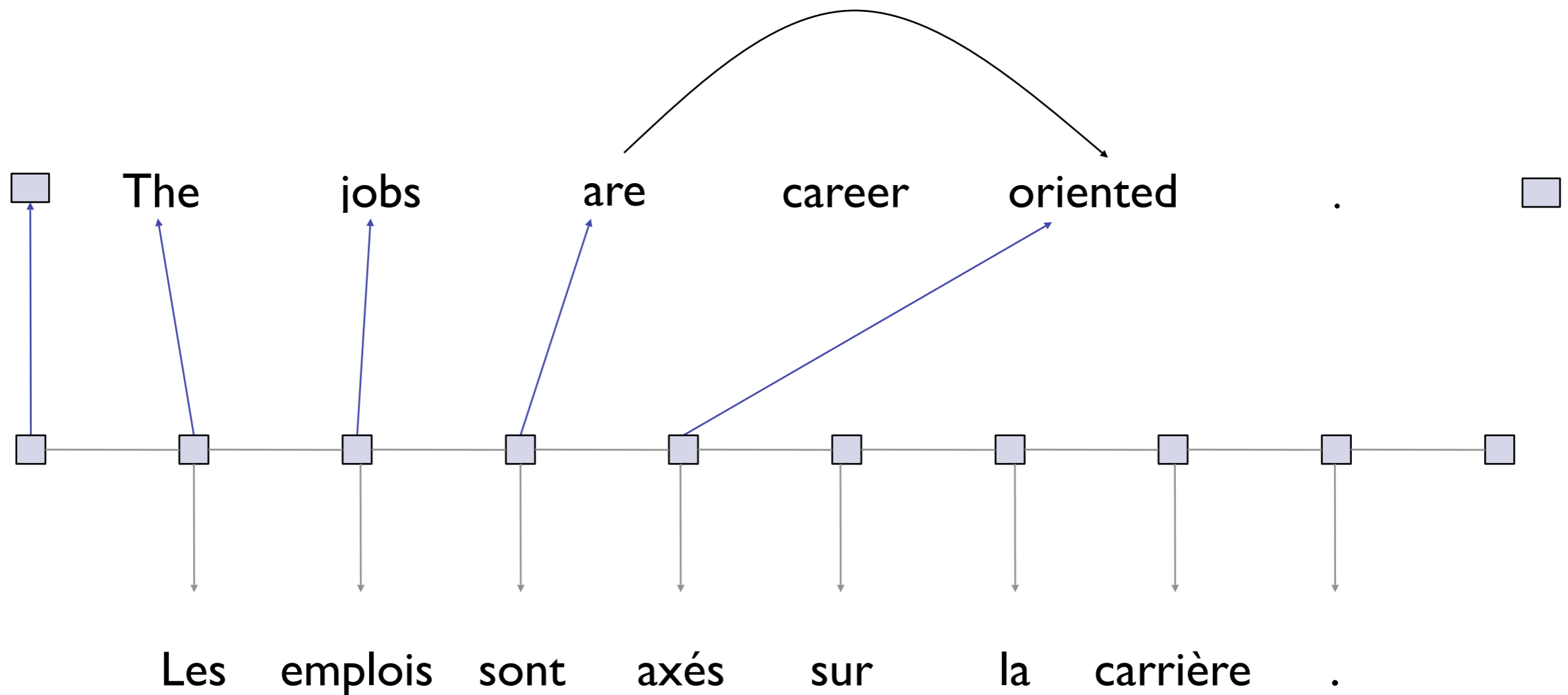
Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$



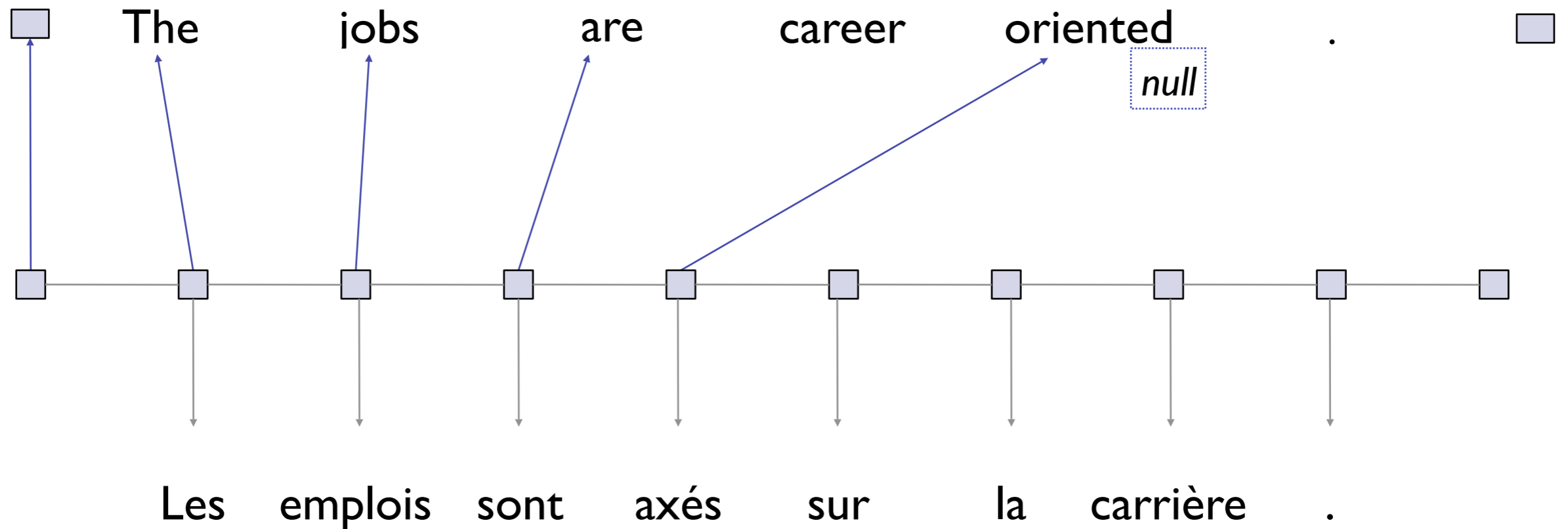
Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j|e_{a_j}) \cdot p(a_j|a_{j-})$$



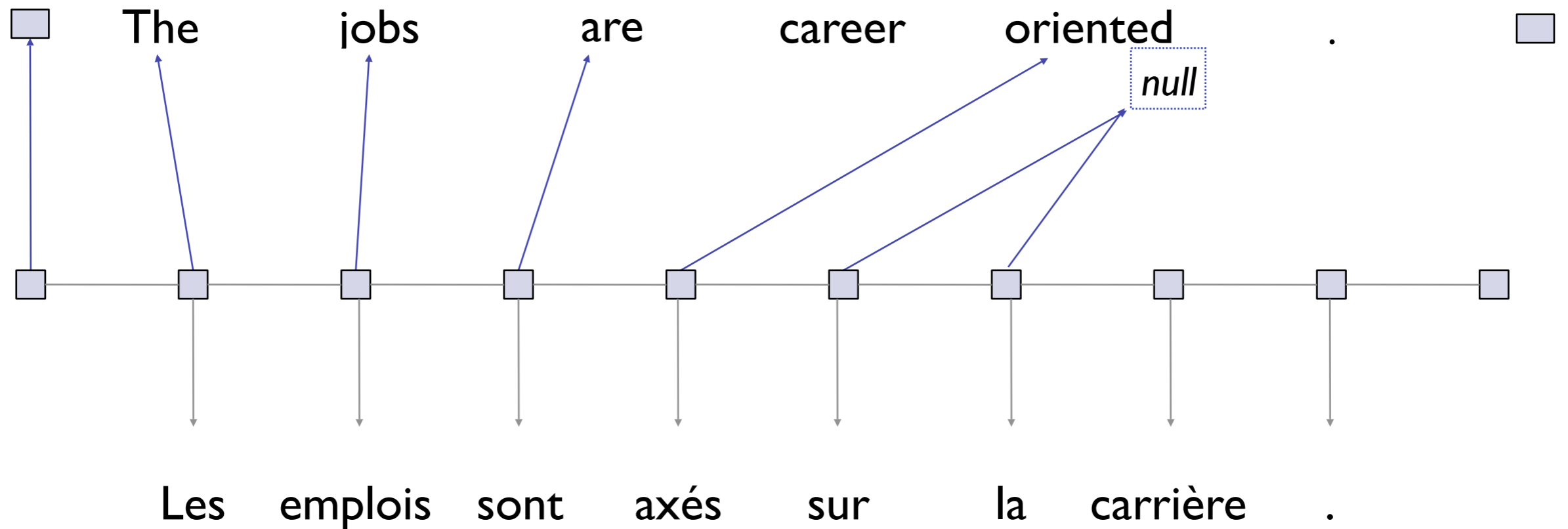
Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$



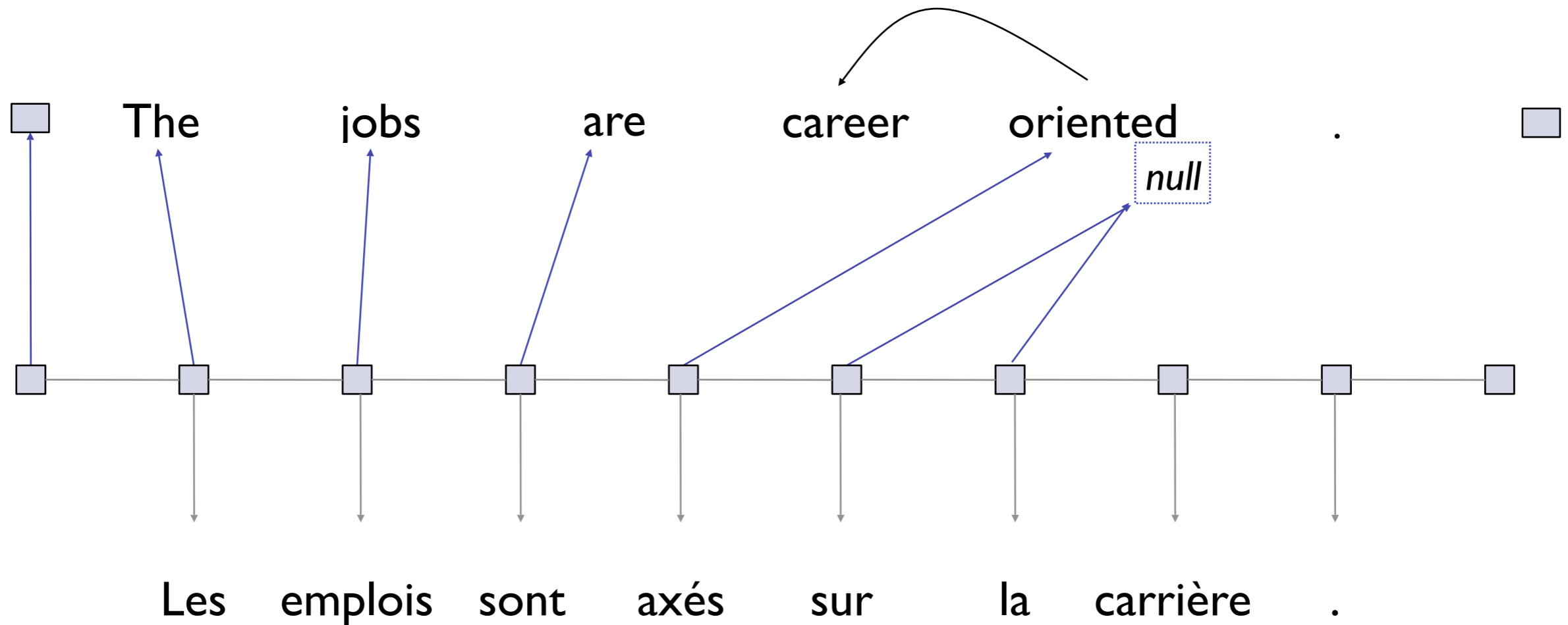
Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$



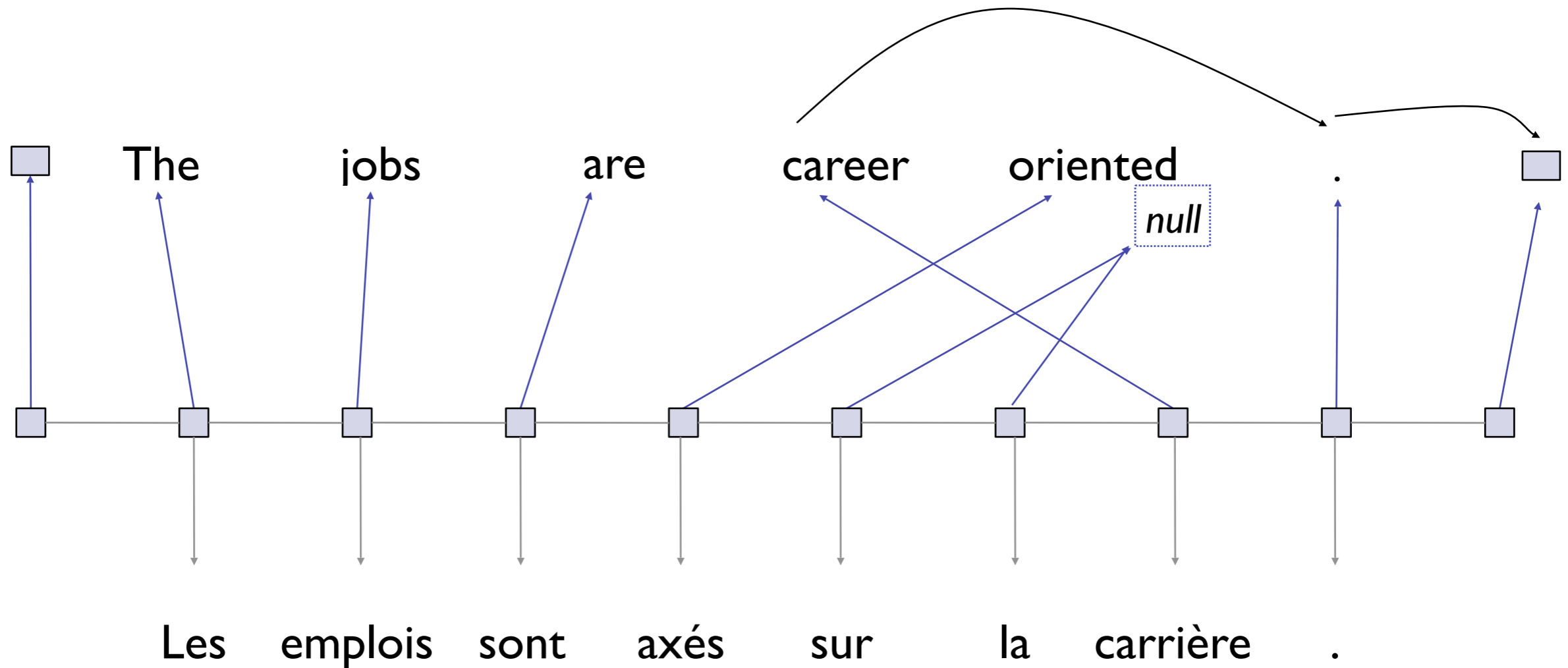
Alignment Errors under the HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$

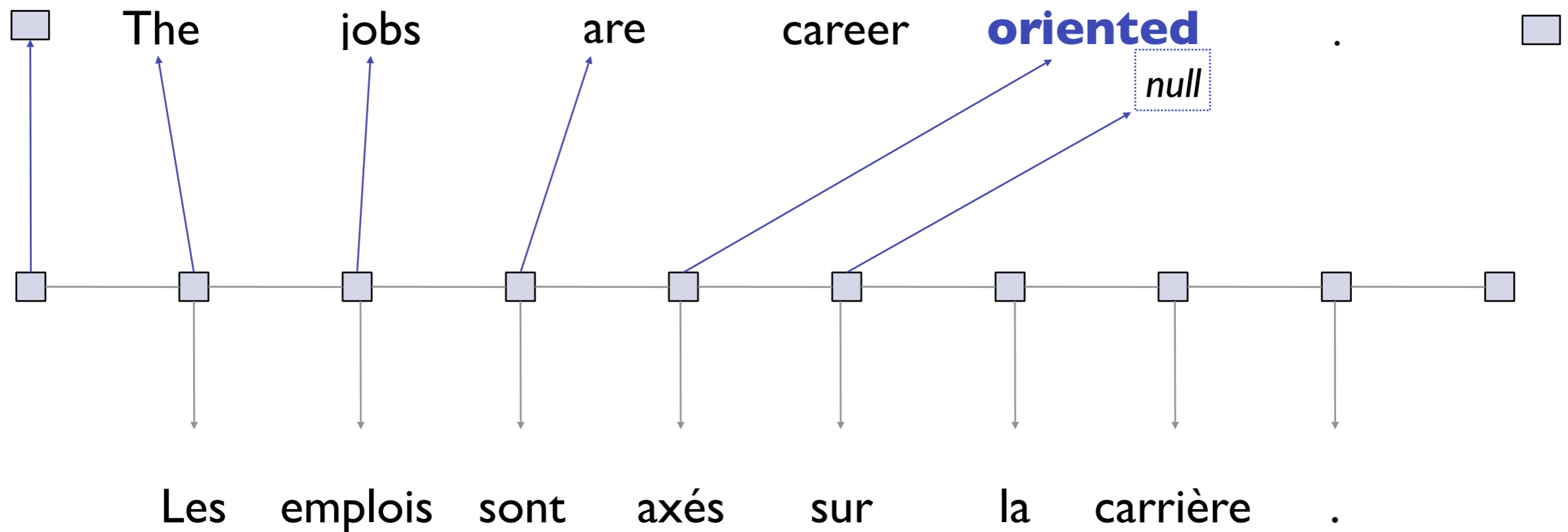


Alignment Errors under the HMM Alignment Model

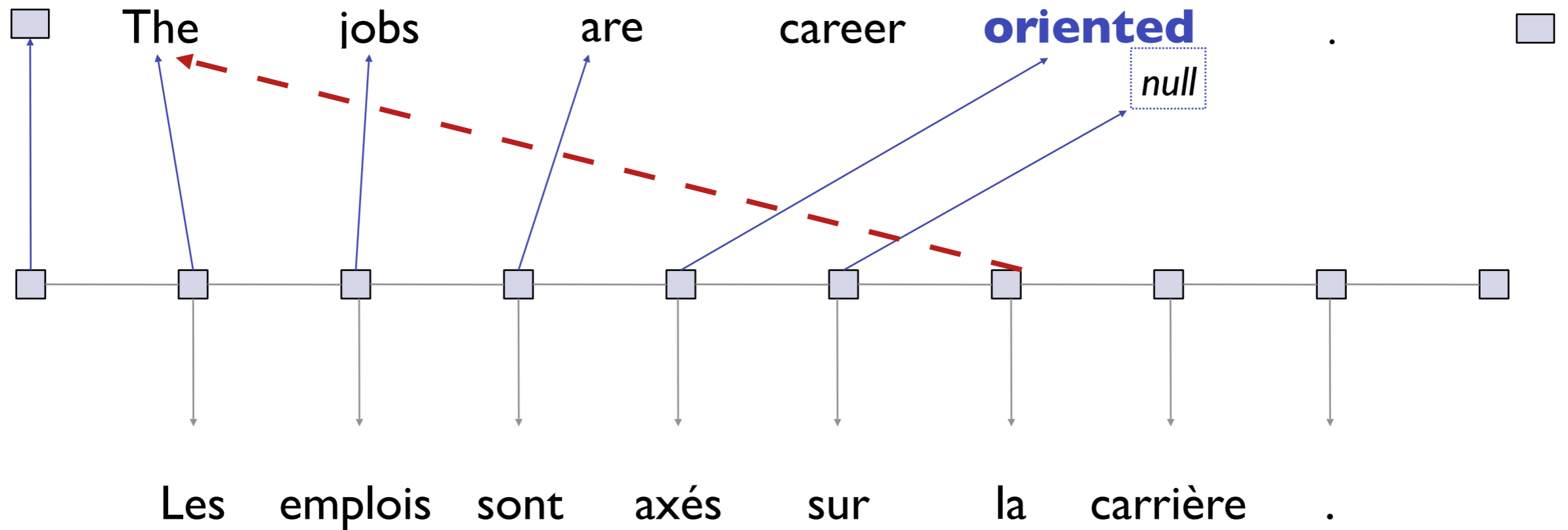
$$p(f, a|e) = \prod_j p(f_j|e_{a_j}) \cdot p(a_j|a_{j-})$$



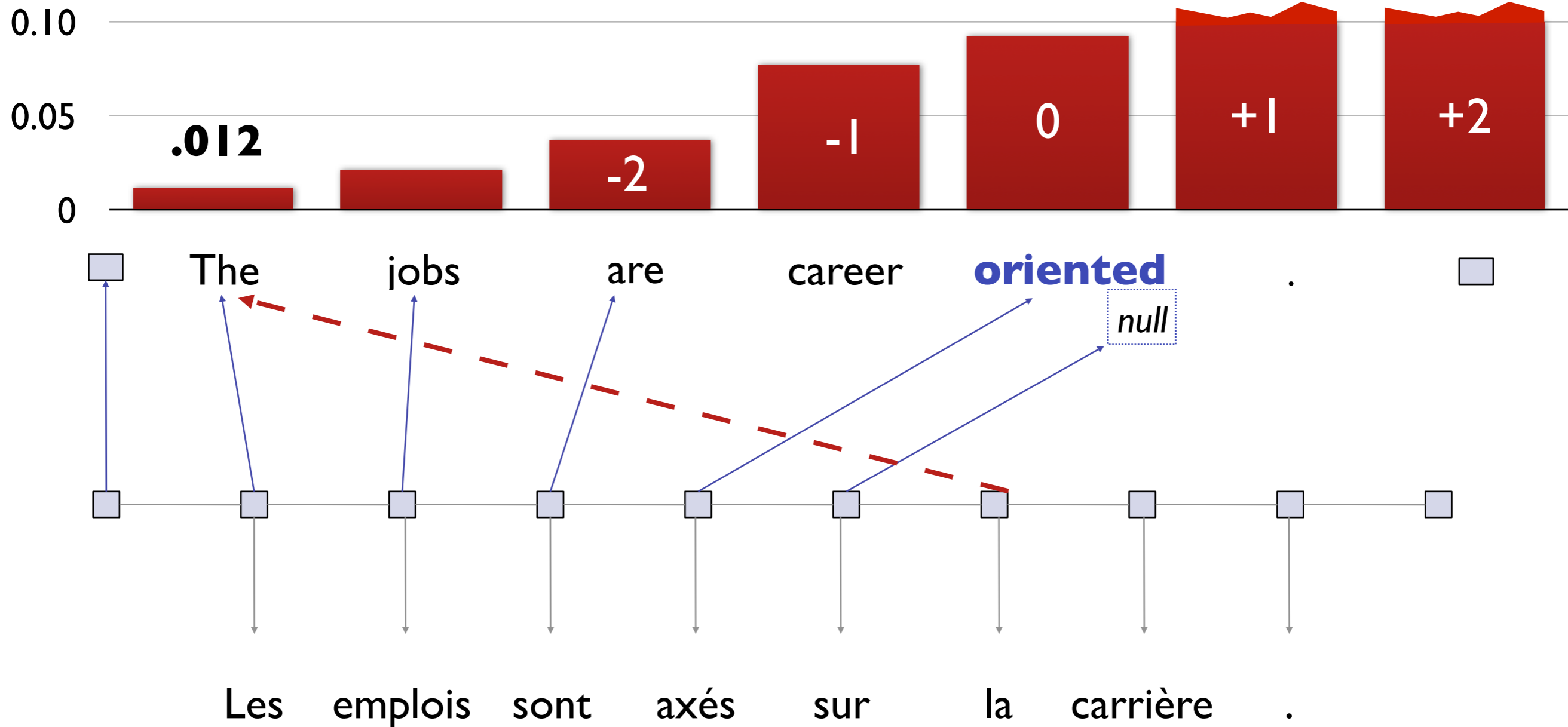
Alignment Errors under the HMM Alignment Model



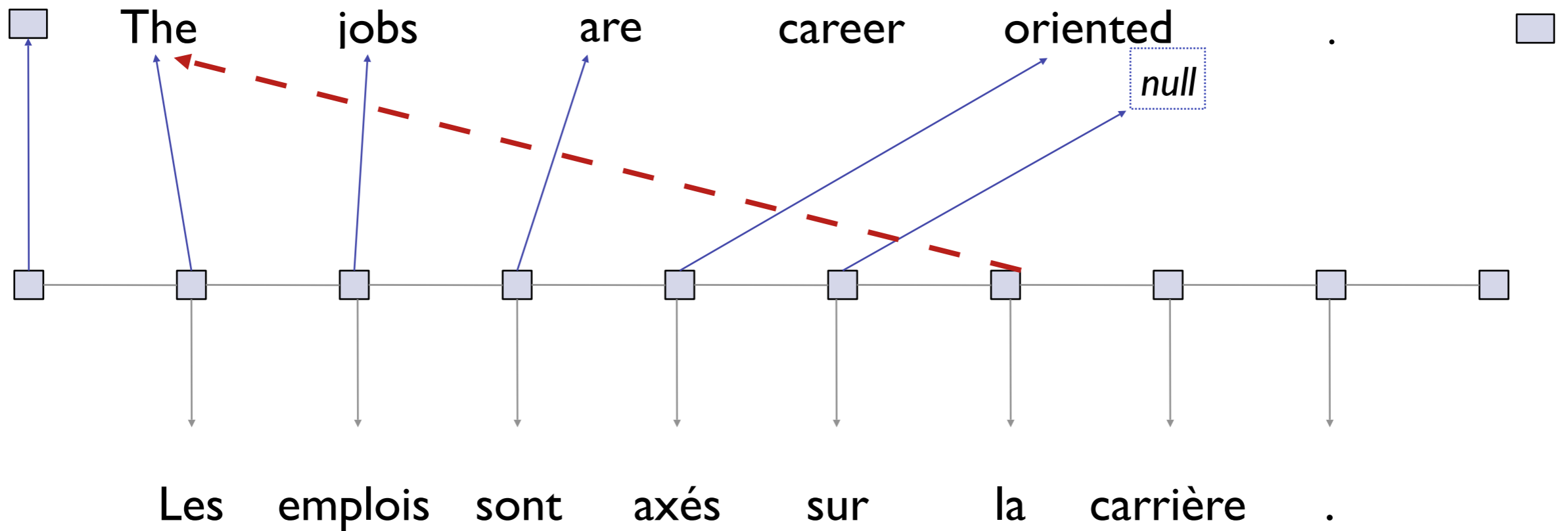
Alignment Errors under the HMM Alignment Model



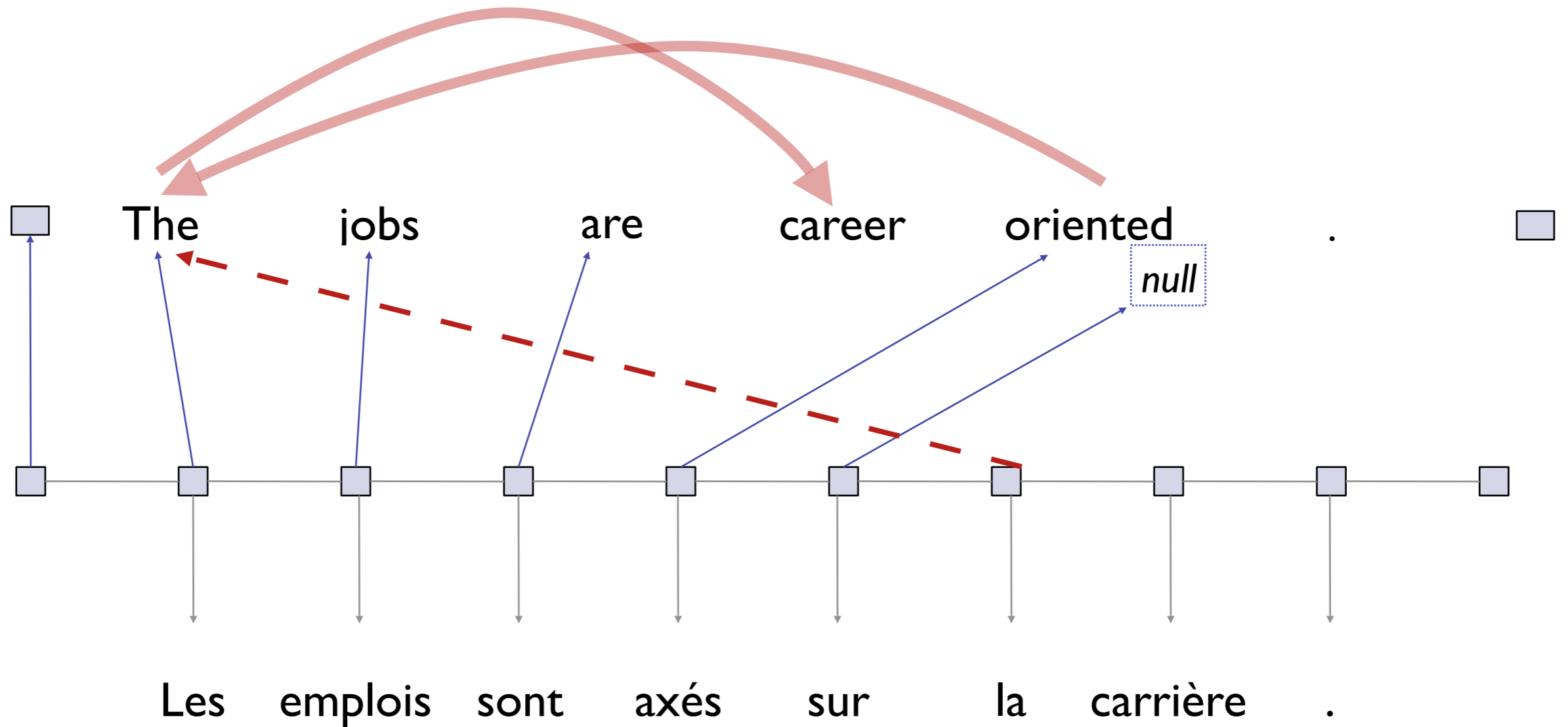
Alignment Errors under the HMM Alignment Model



Syntactic HMM Alignment Model

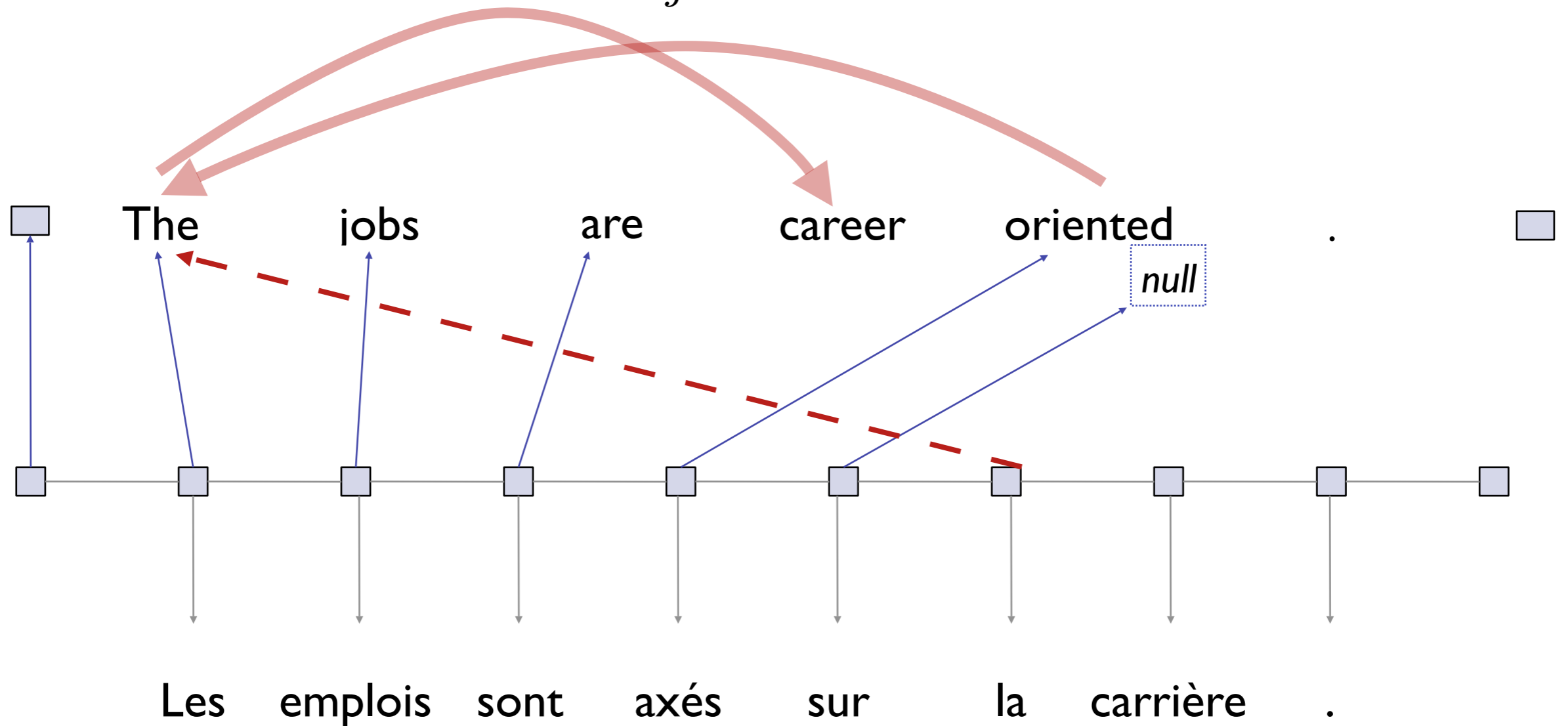


Syntactic HMM Alignment Model



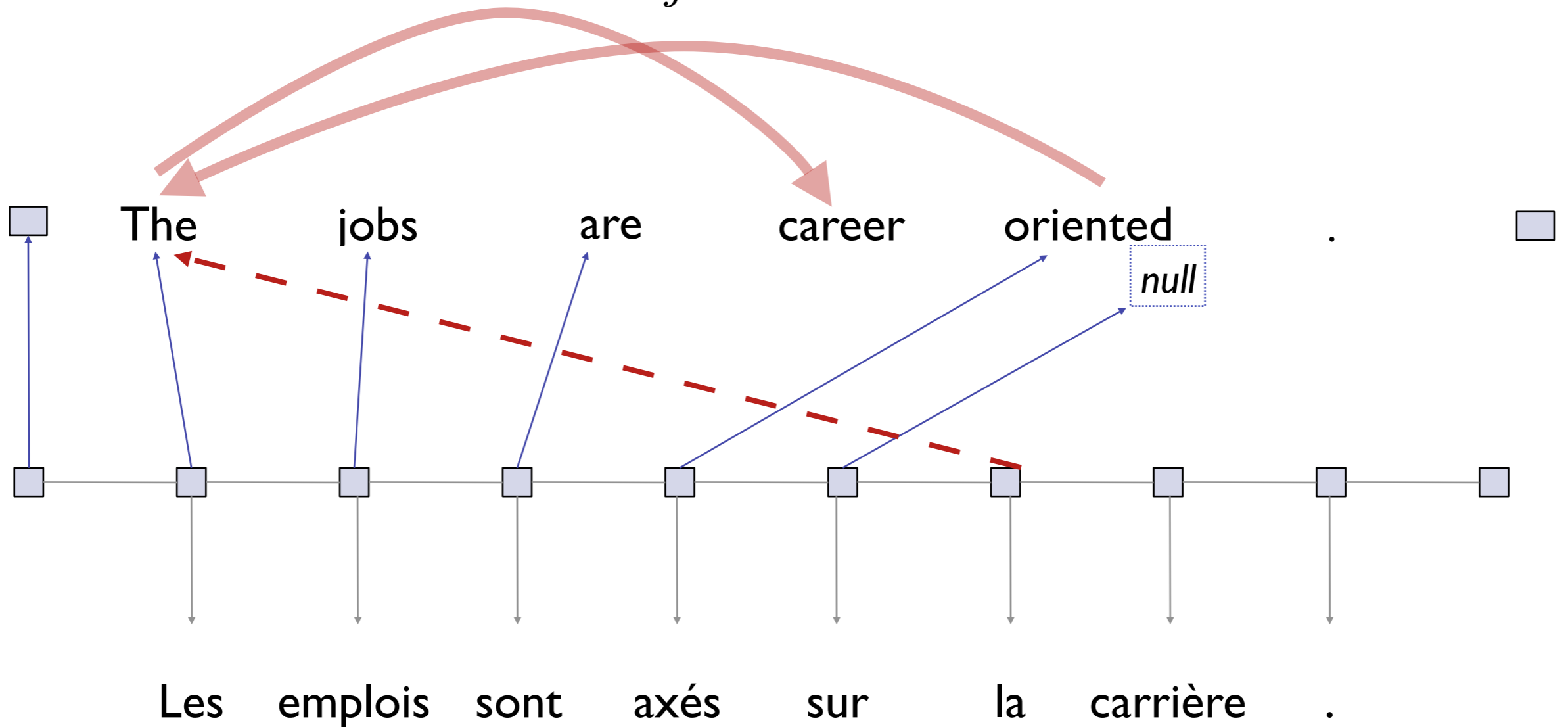
Syntactic HMM Alignment Model

$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$

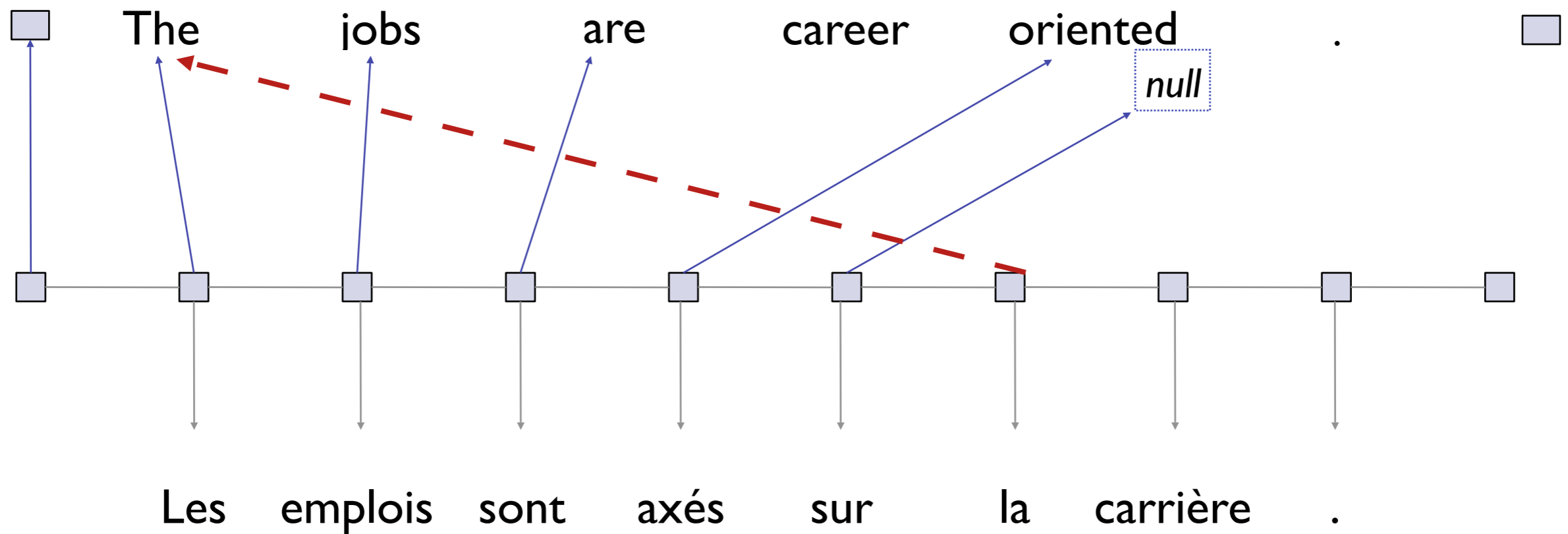


Syntactic HMM Alignment Model

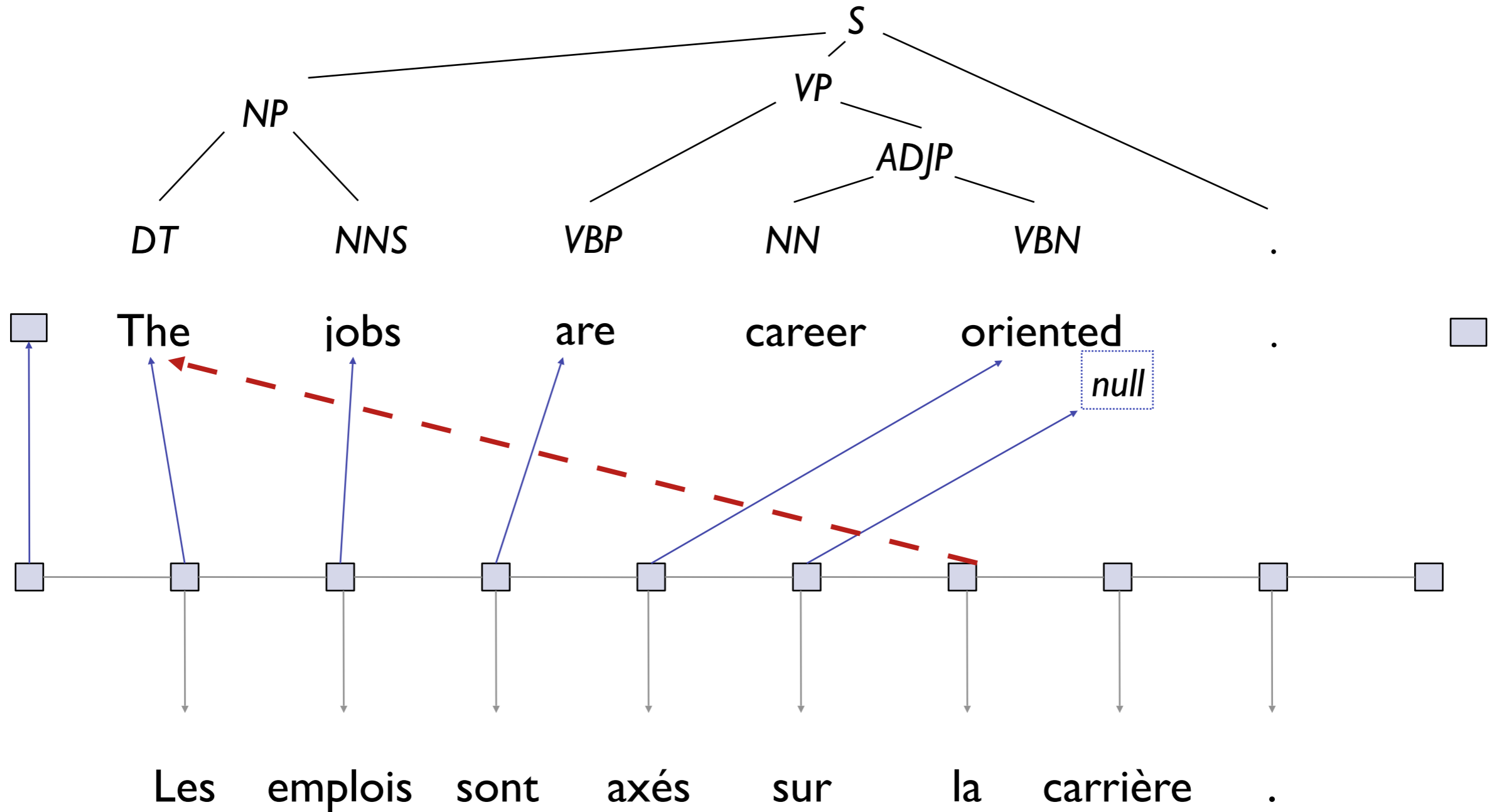
$$p(f, a|e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-}, t)$$



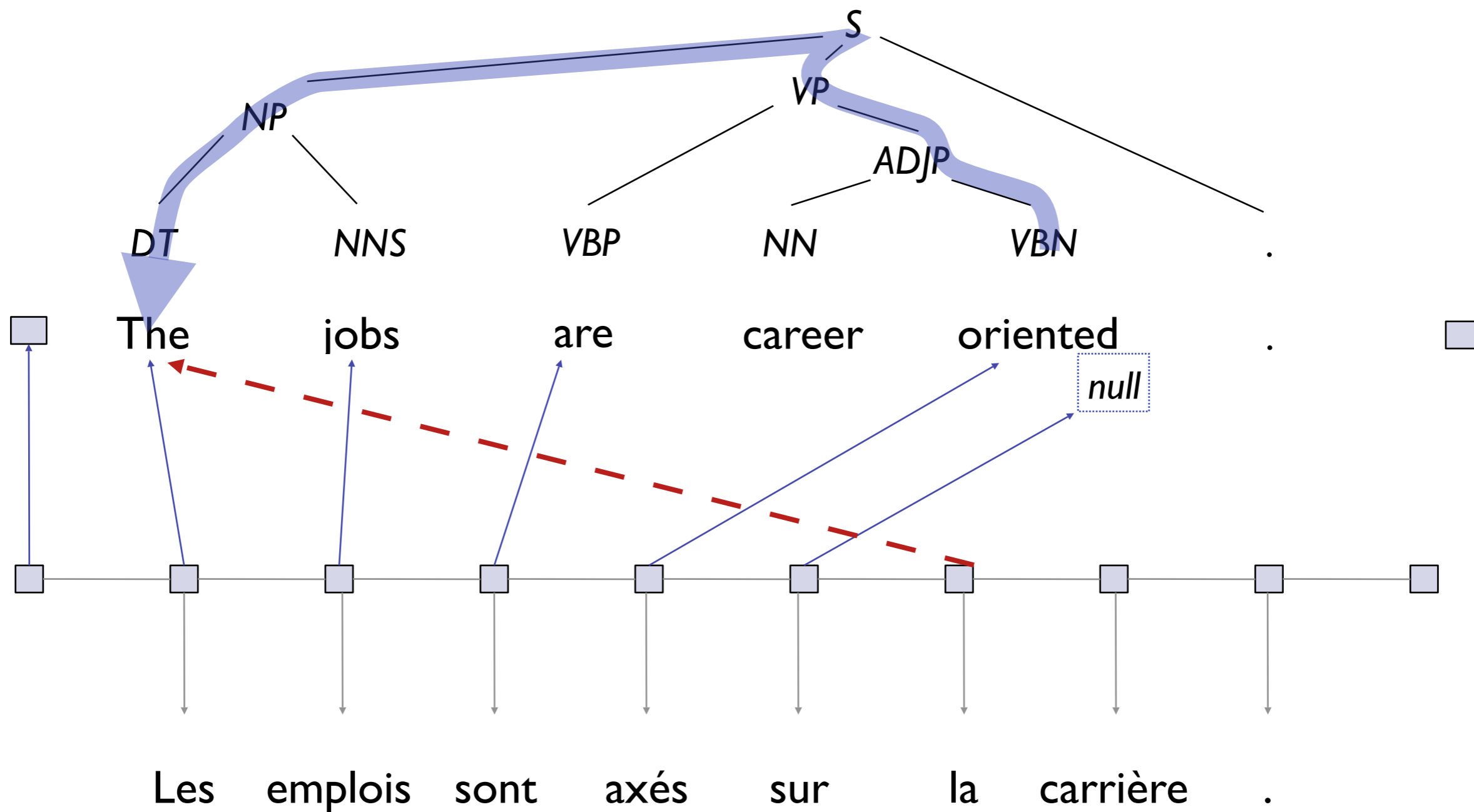
Syntactic HMM Alignment Model



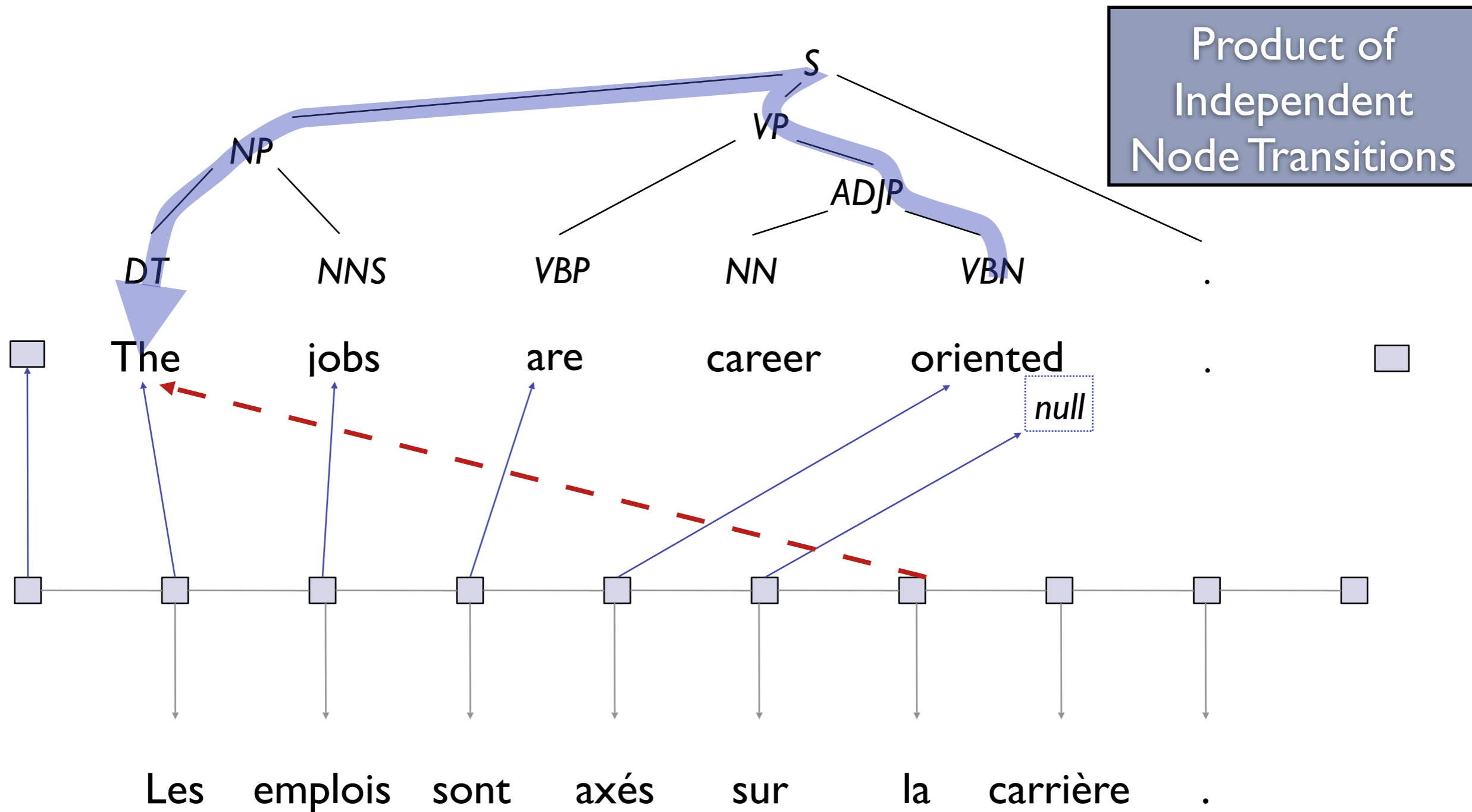
Syntactic HMM Alignment Model



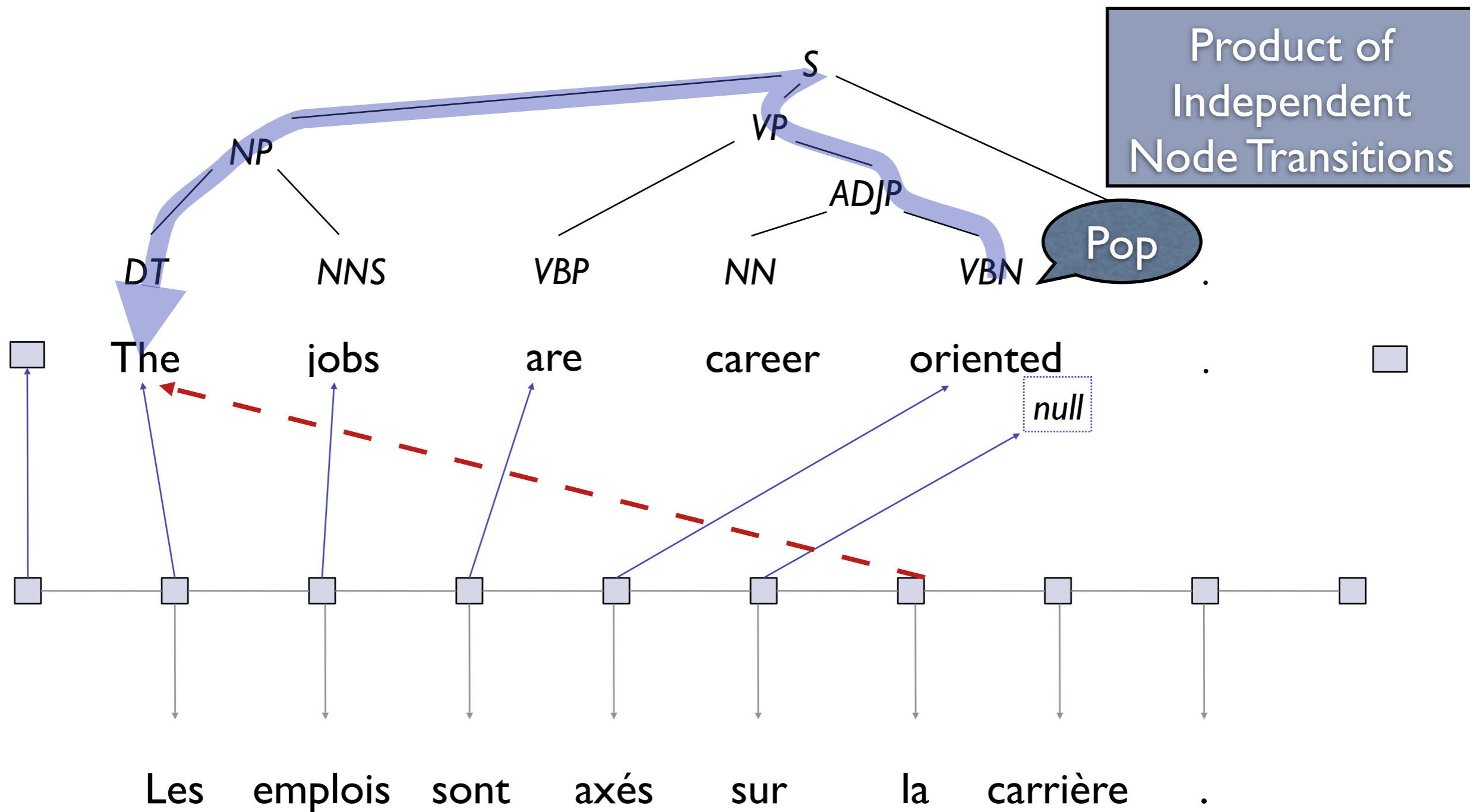
Syntactic HMM Alignment Model



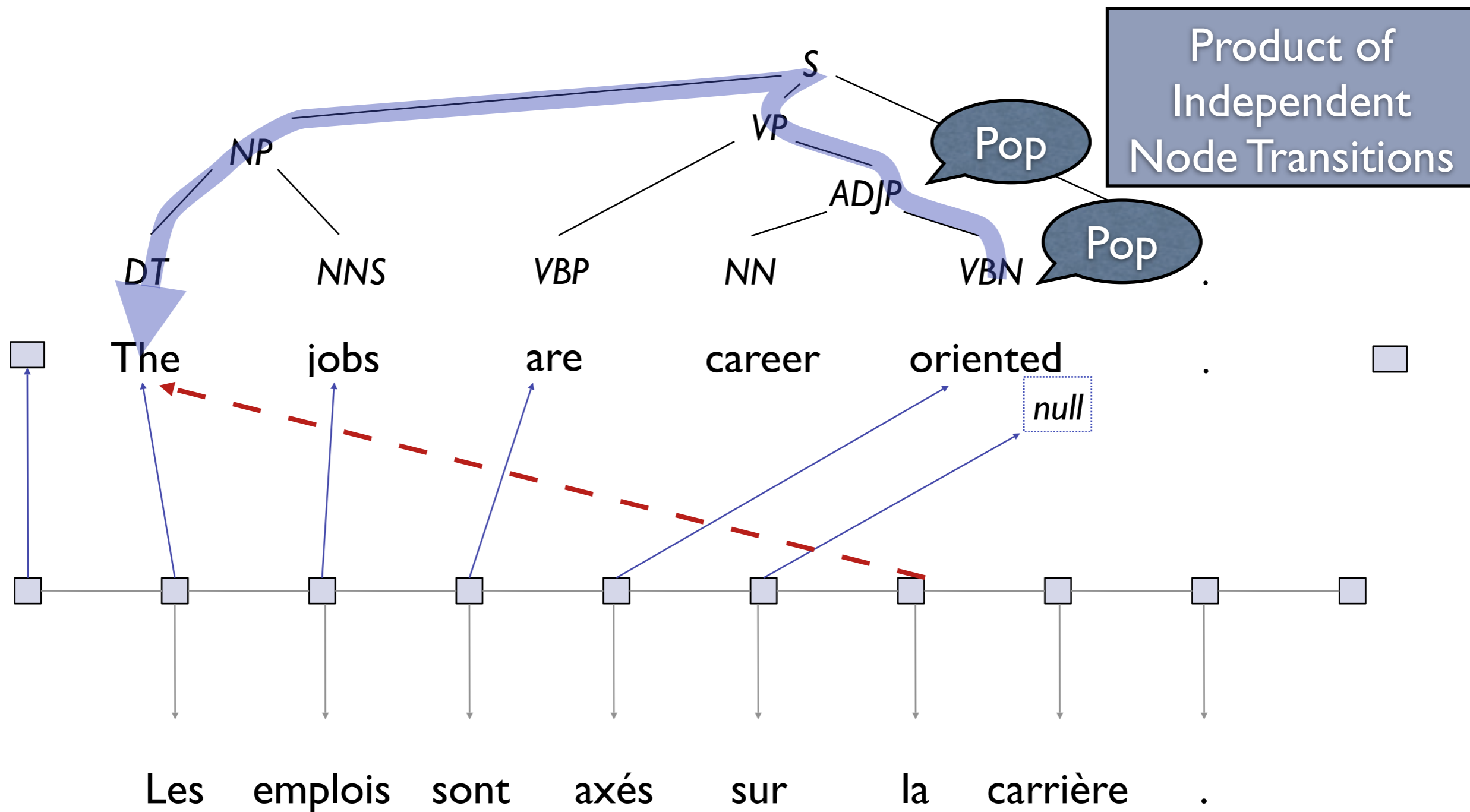
Syntactic HMM Alignment Model



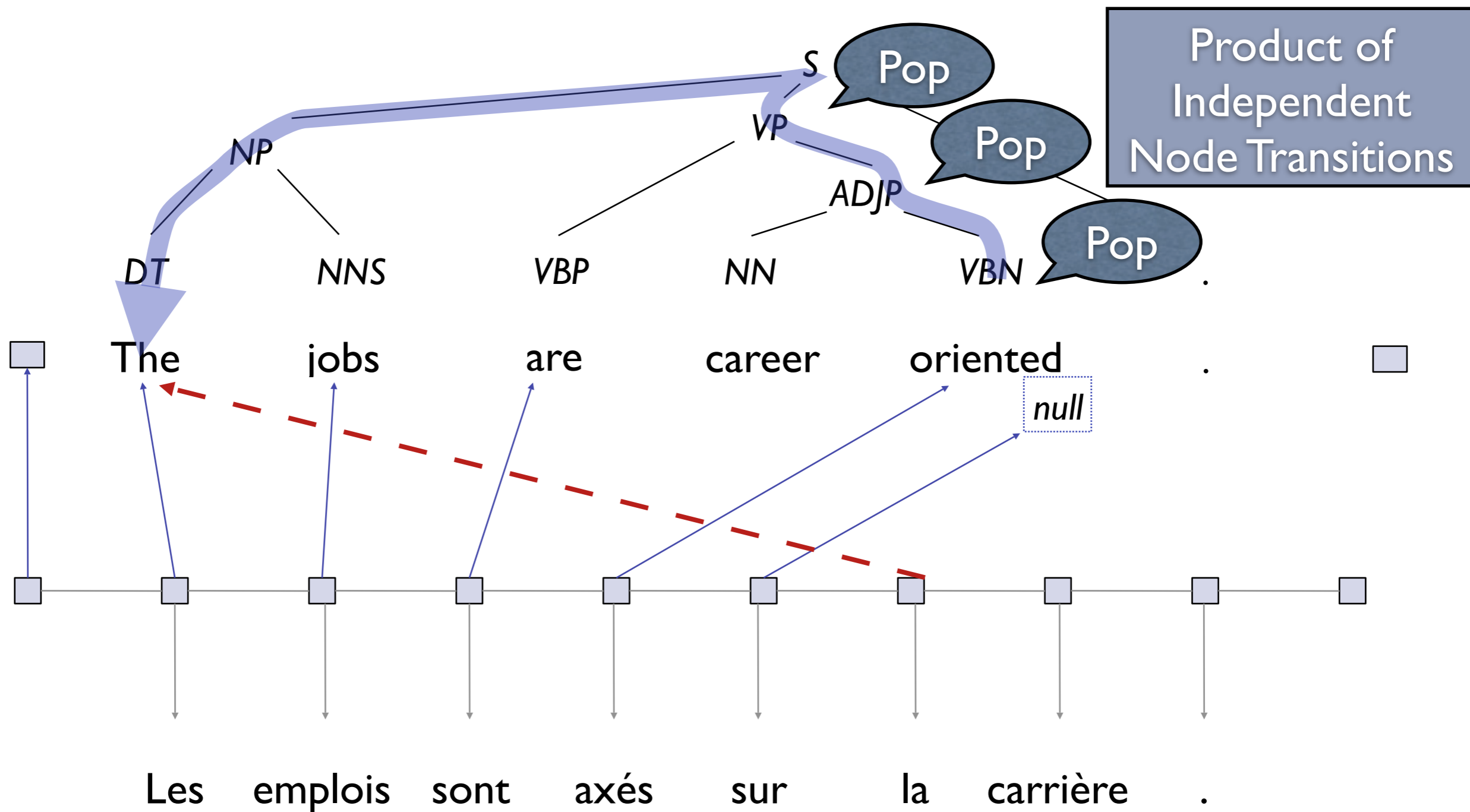
Syntactic HMM Alignment Model



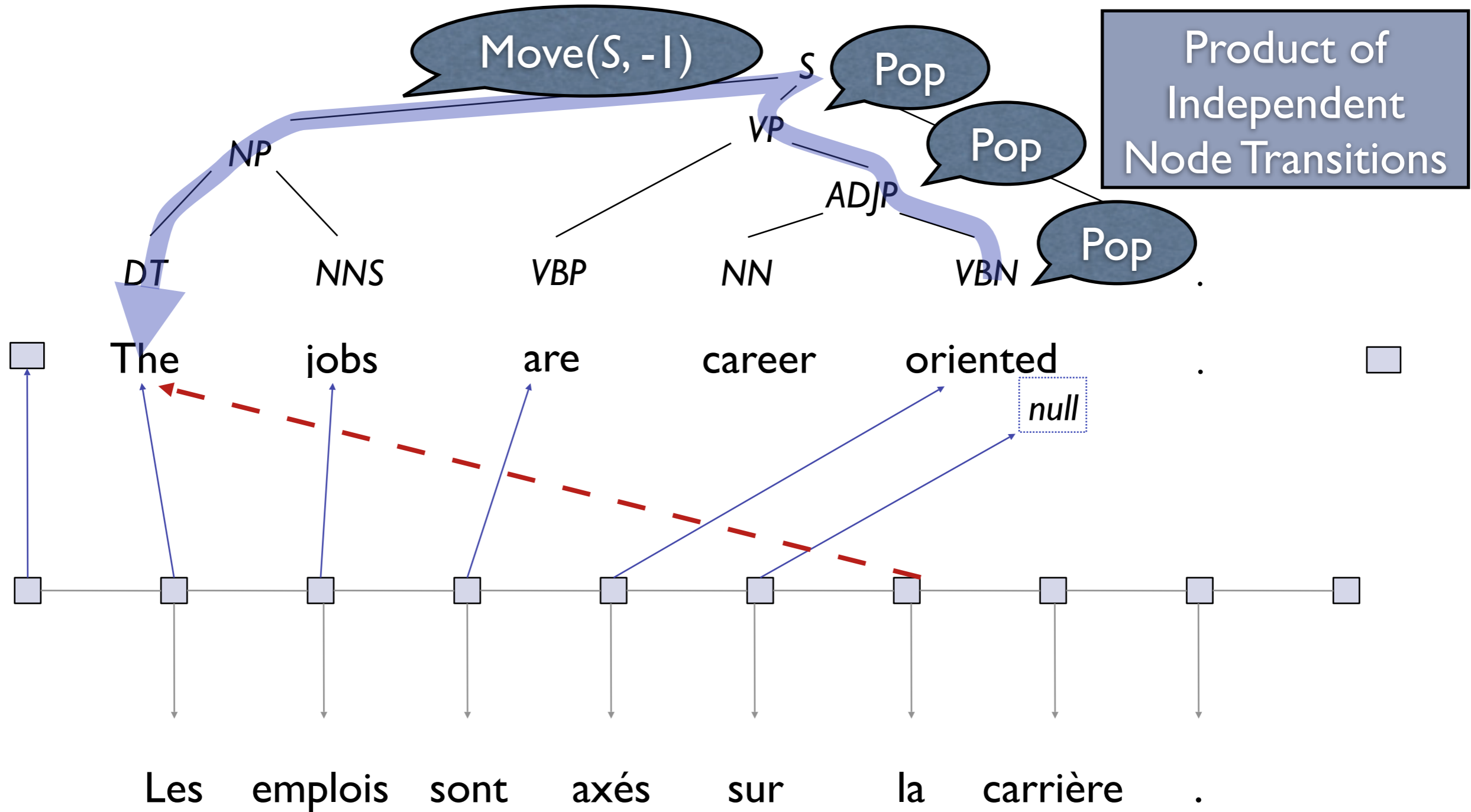
Syntactic HMM Alignment Model



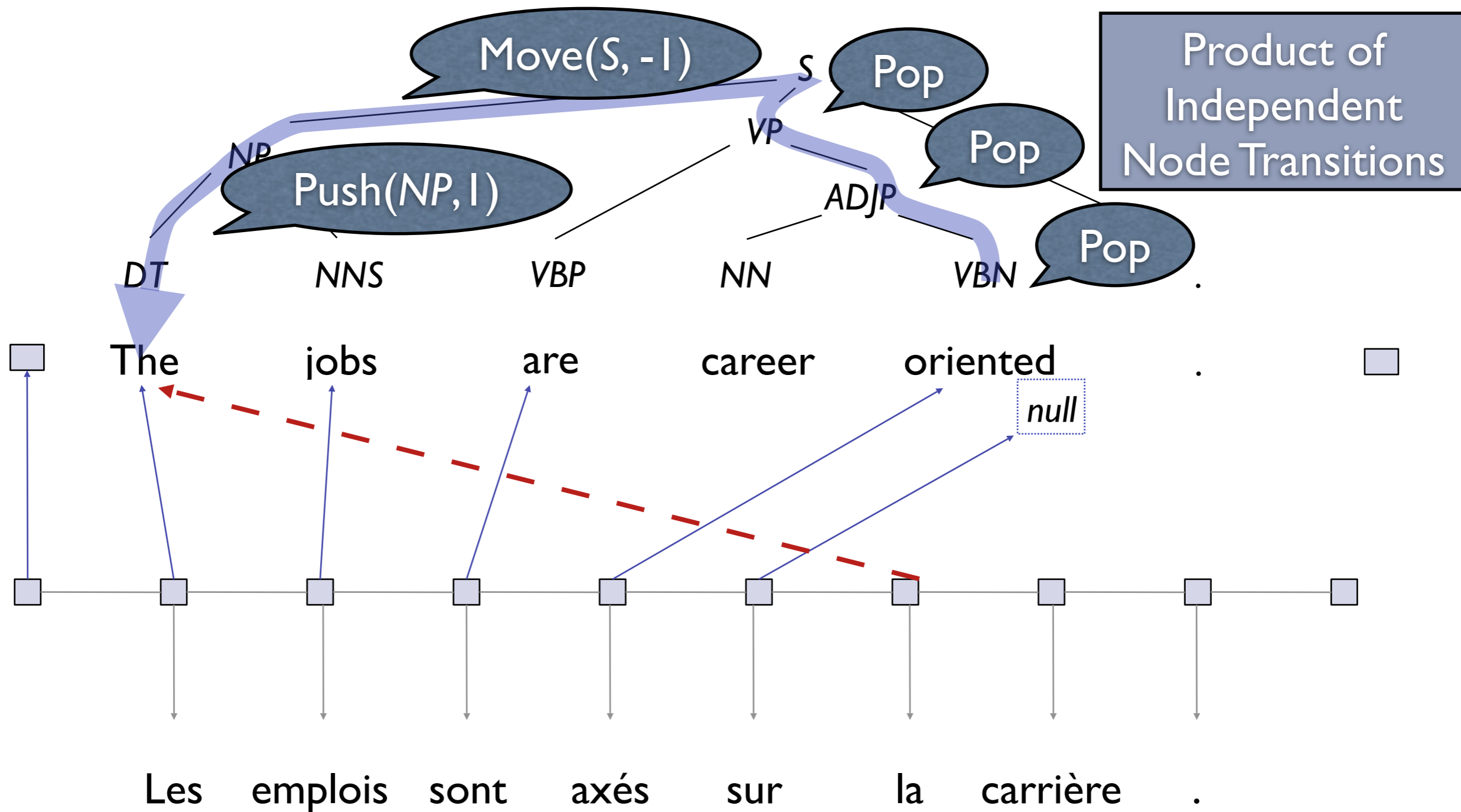
Syntactic HMM Alignment Model



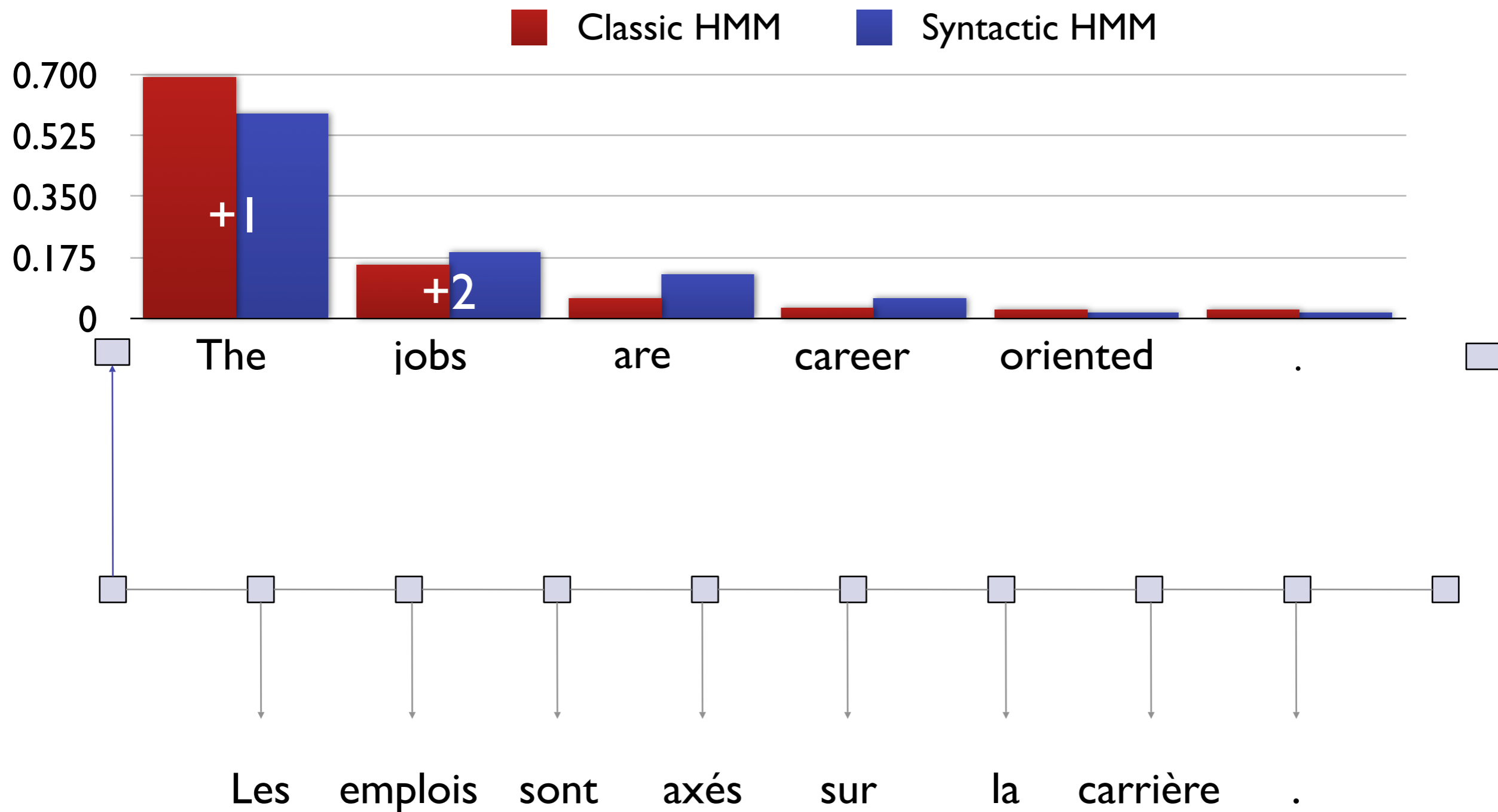
Syntactic HMM Alignment Model



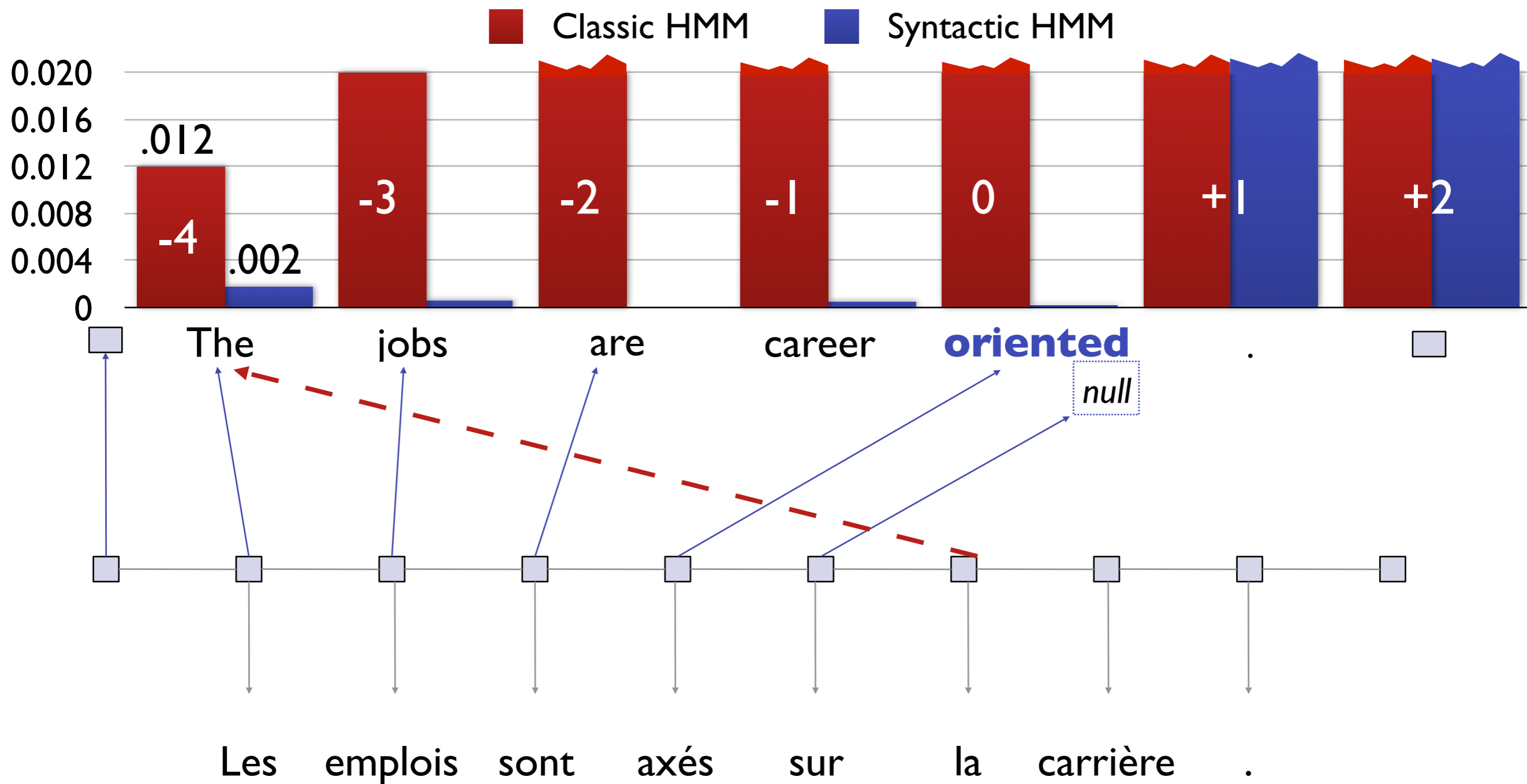
Syntactic HMM Alignment Model



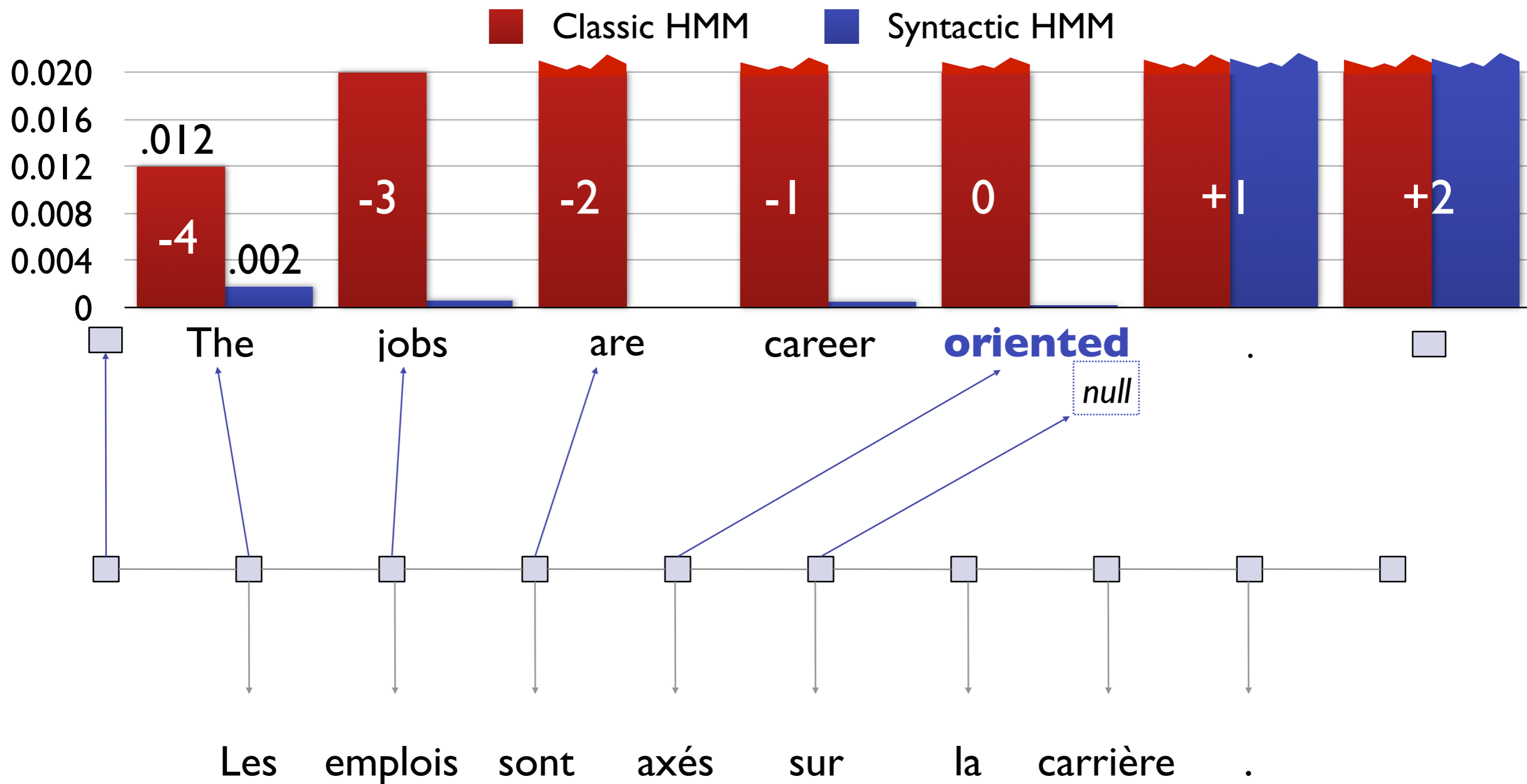
Syntactic HMM Alignment Model



Syntactic HMM Alignment Model



Syntactic HMM Alignment Model





Evaluation: Alignment Error Rate (AER)



Evaluation: Alignment Error Rate (AER)

Test Conditions

- Chinese-English from MT-Eval 02 test set
- 100k training sentences from FBIS
- Initialized with agreement training for Model 1 (Liang et al., 06)



Evaluation: Alignment Error Rate (AER)

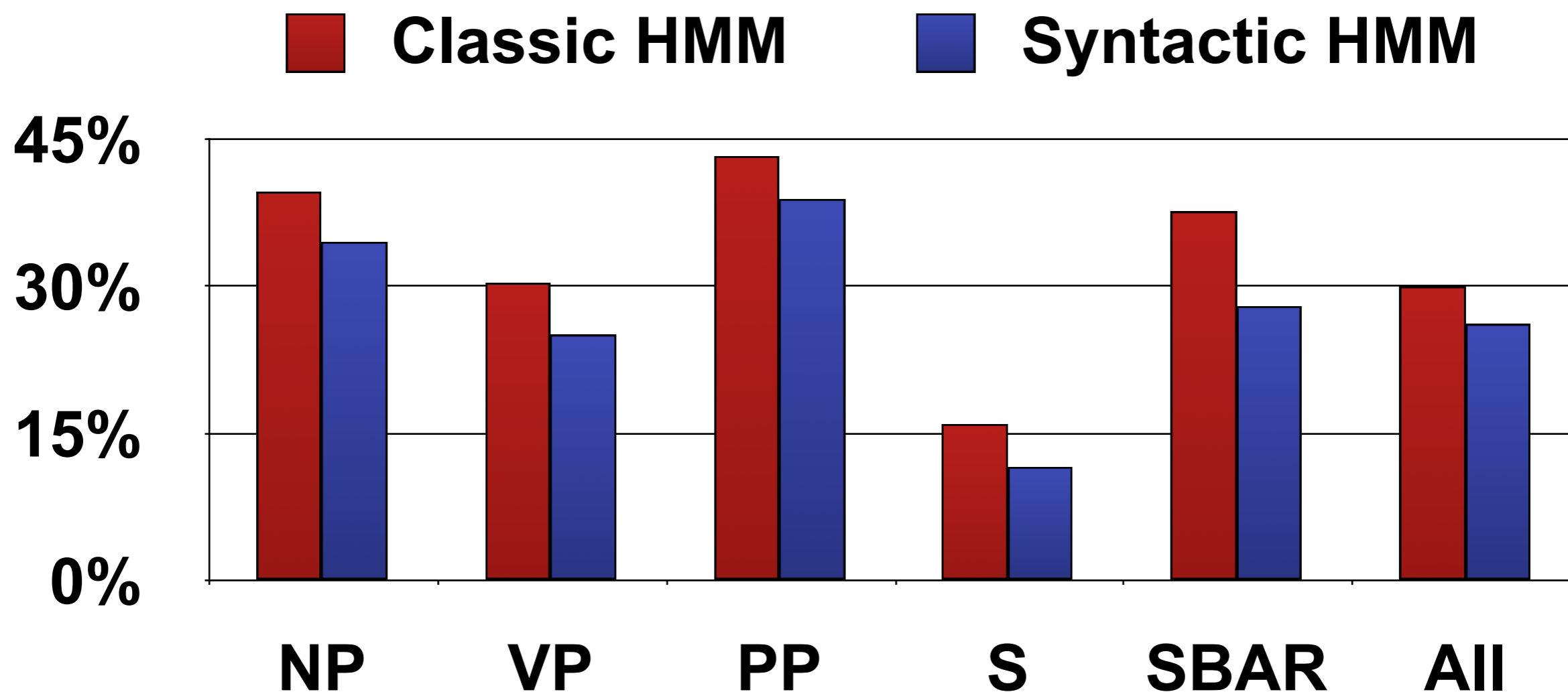
Test Conditions

- Chinese-English from MT-Eval 02 test set
- 100k training sentences from FBIS
- Initialized with agreement training for Model 1 (Liang et al., 06)

Results

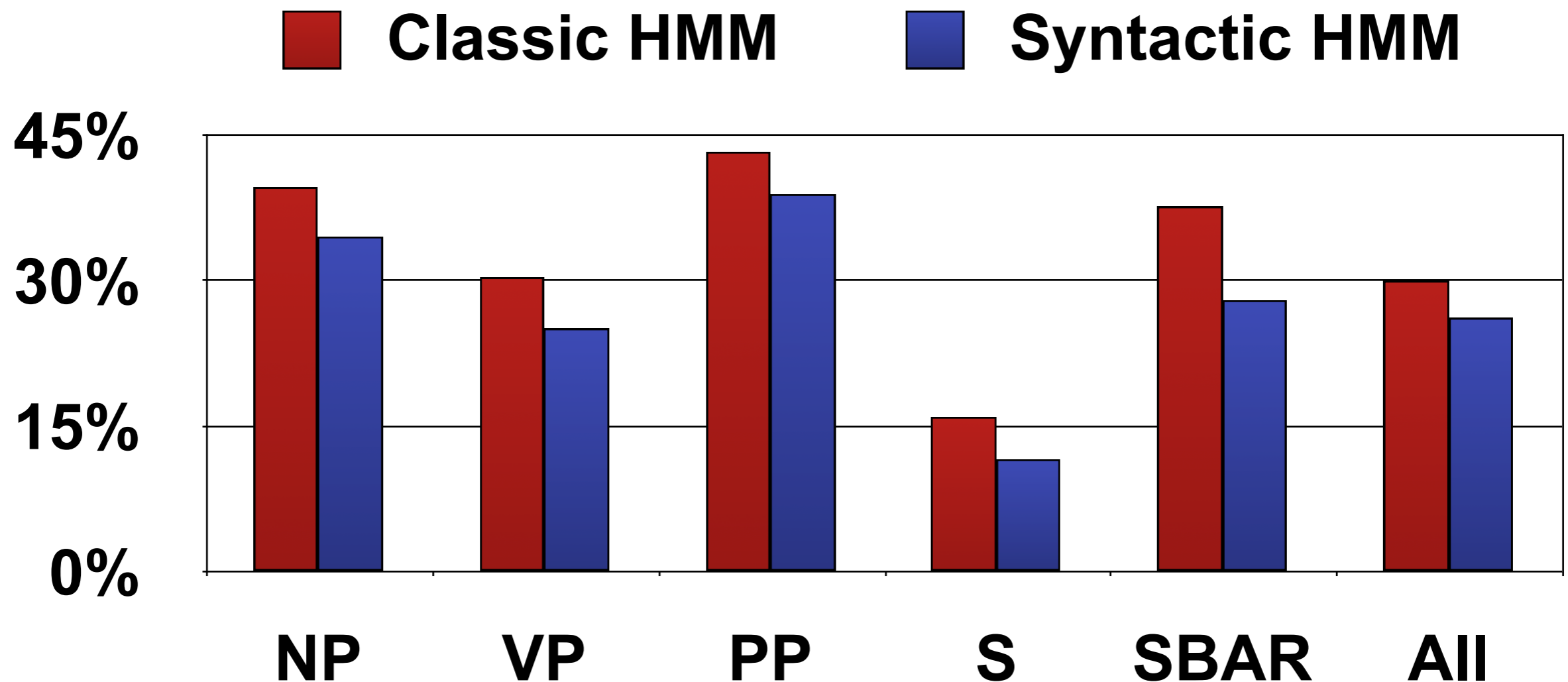
	Precision	Recall	AER
Classic HMM	81.6	78.8	19.8
Syntactic HMM	82.2	76.8	20.5
GIZA++	61.9	82.6	29.7

Evaluation: Unproductive Constituent Rates



Evaluation: Unproductive Constituent Rates

The Syntactic HMM Reduces the Frequency of Unproductive *Interior* Nodes by 13%



Decoding Heuristic: Competitive Thresholding

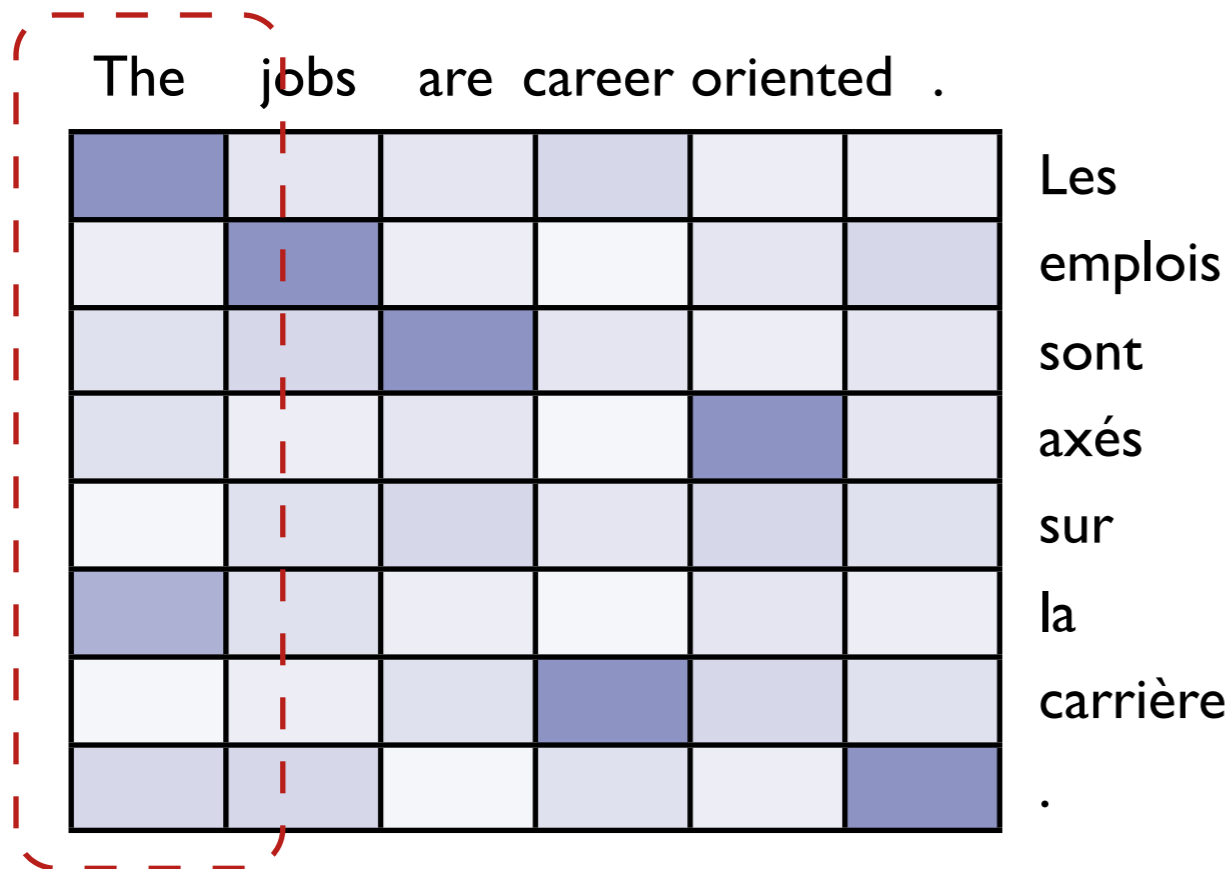
Only the maximum posterior in each row or column and its neighbors can be included in the alignment

The jobs are career oriented .

						Les
						emplois
						sont
						axés
						sur
						la
						carrière
						.

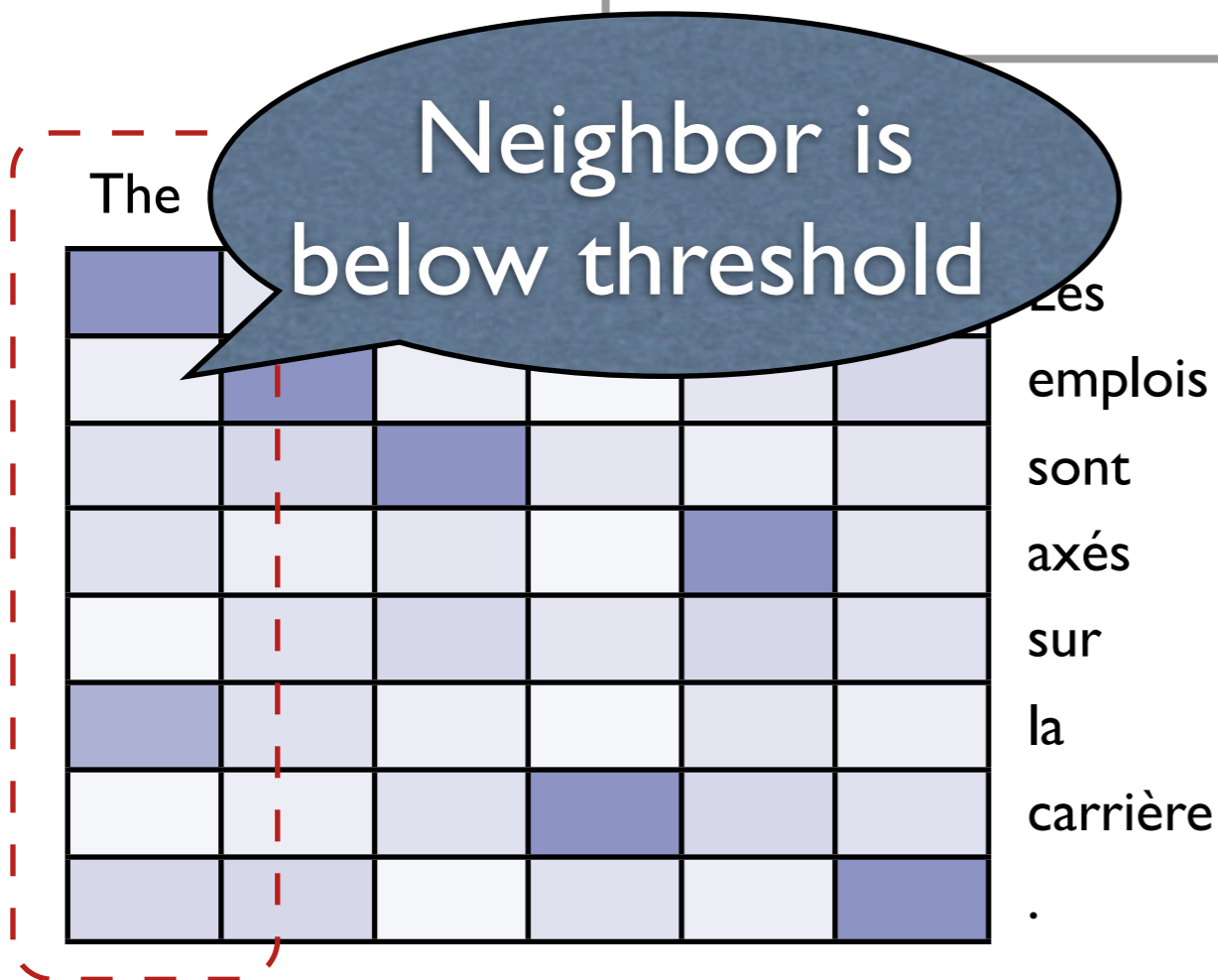
Decoding Heuristic: Competitive Thresholding

Only the maximum posterior in each row or column and its neighbors can be included in the alignment



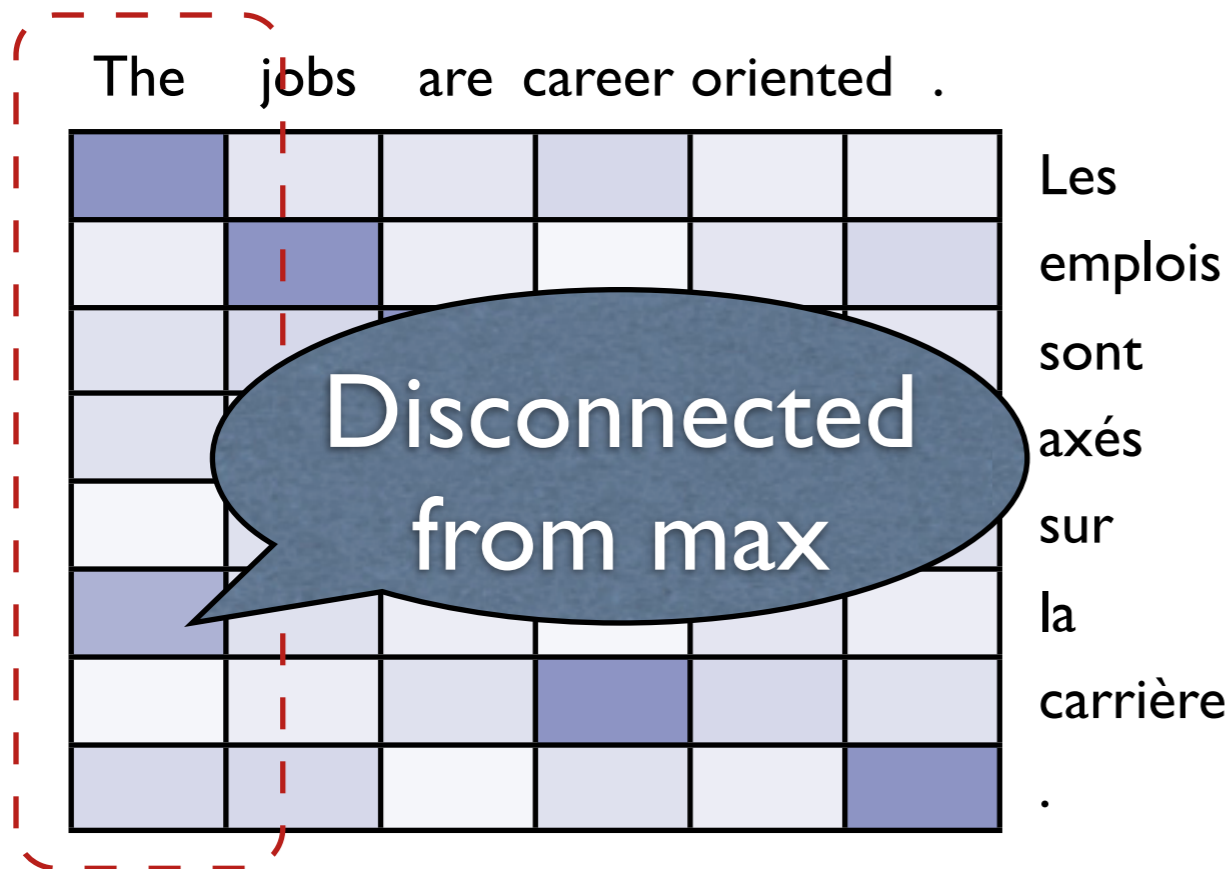
Decoding Heuristic: Competitive Thresholding

Only the maximum posterior in each row or column and its neighbors can be included in the alignment



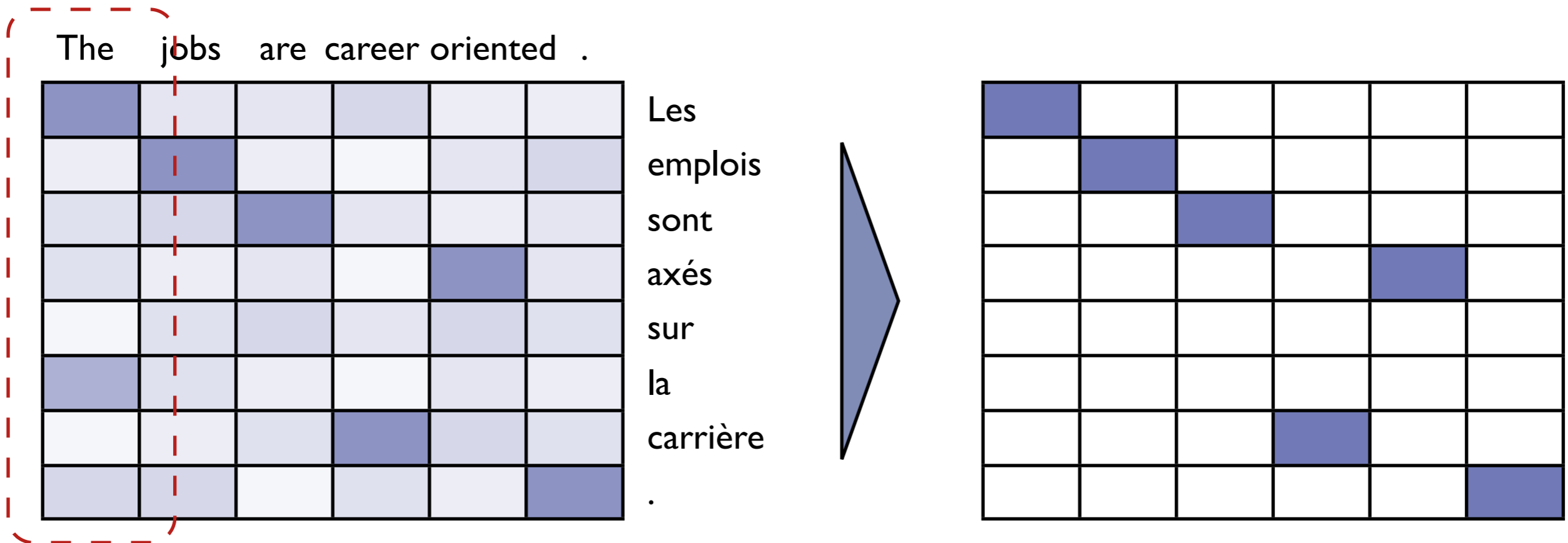
Decoding Heuristic: Competitive Thresholding

Only the maximum posterior in each row or column and its neighbors can be included in the alignment



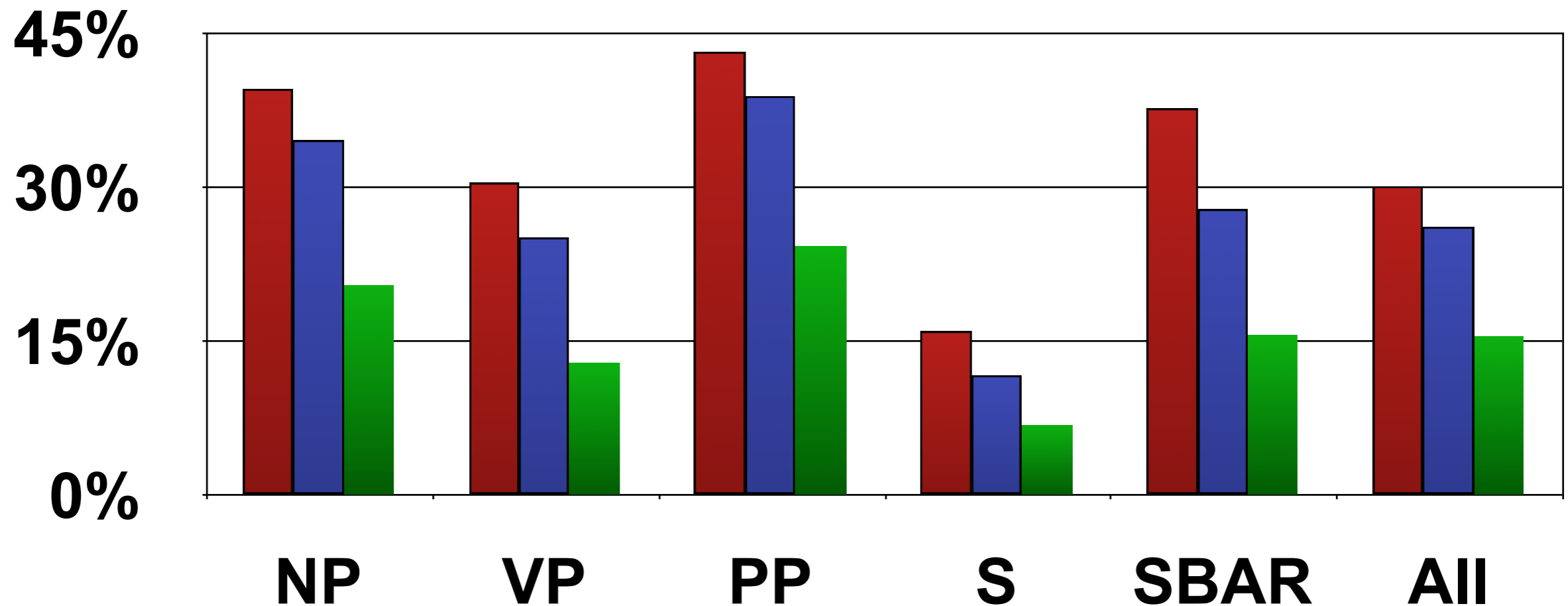
Decoding Heuristic: Competitive Thresholding

Only the maximum posterior in each row or column and its neighbors can be included in the alignment



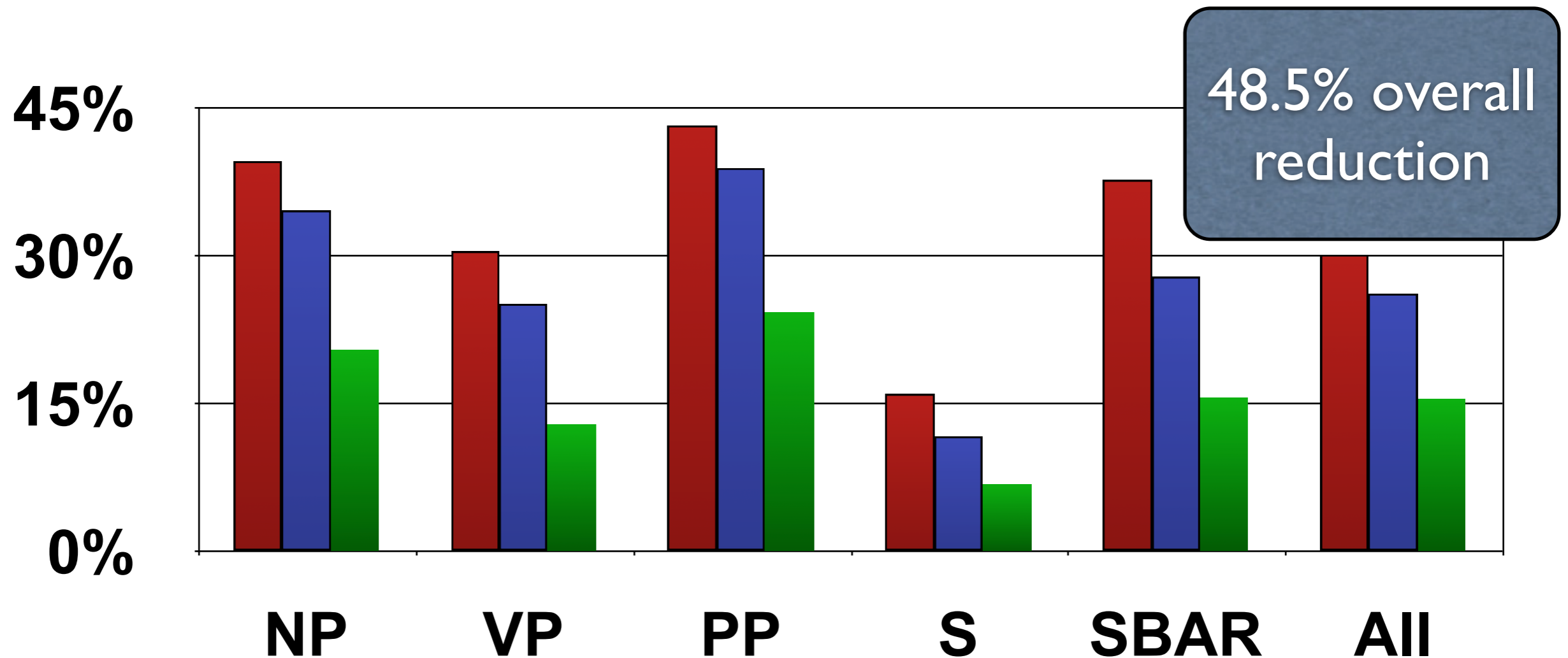
Evaluation: Unproductive Constituent Rates

- Classic HMM**
- Syntactic HMM**
- Syntactic HMM with Competitive Thresholding**



Evaluation: Unproductive Constituent Rates

- Classic HMM**
- Syntactic HMM**
- Syntactic HMM with Competitive Thresholding**





Evaluation: Quantity of Rules Extracted

Rules extracted per sentence



Evaluation: Quantity of Rules Extracted

Rules extracted per sentence

Syntactic HMM + CT

Syntactic HMM

Classic HMM



Evaluation: Quantity of Rules Extracted

Rules extracted per sentence

Syntactic HMM + CT

Syntactic HMM

Classic HMM

65

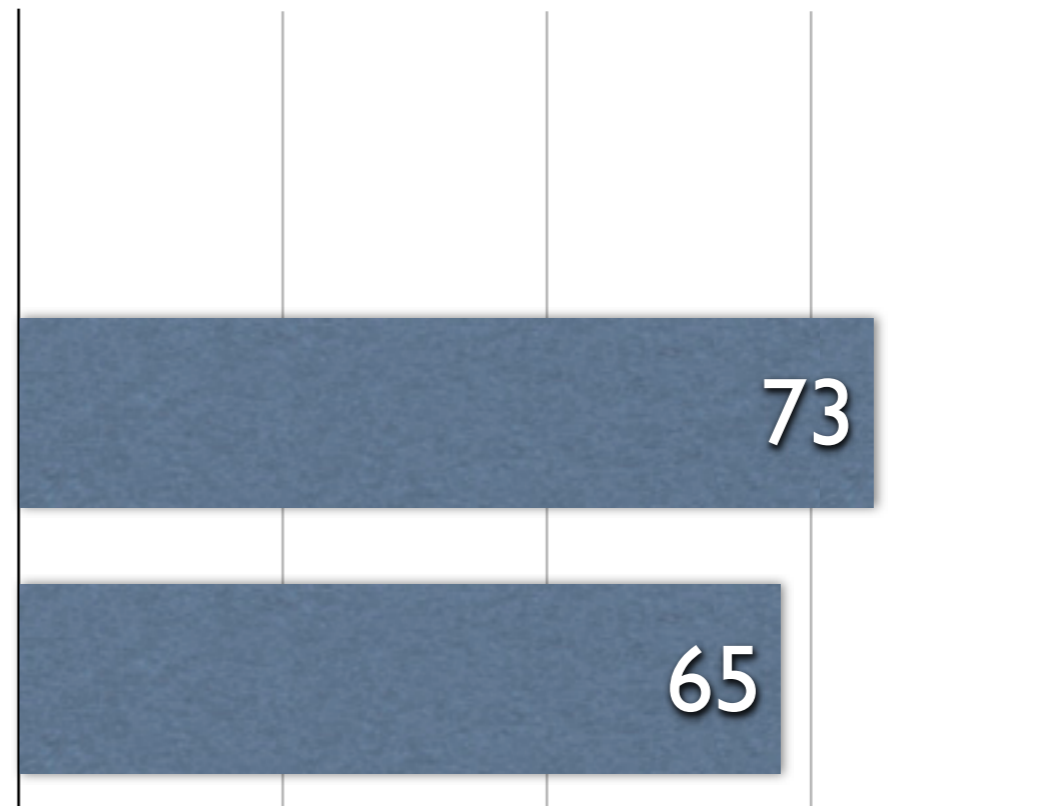
Evaluation: Quantity of Rules Extracted

Rules extracted per sentence

Syntactic HMM + CT

Syntactic HMM

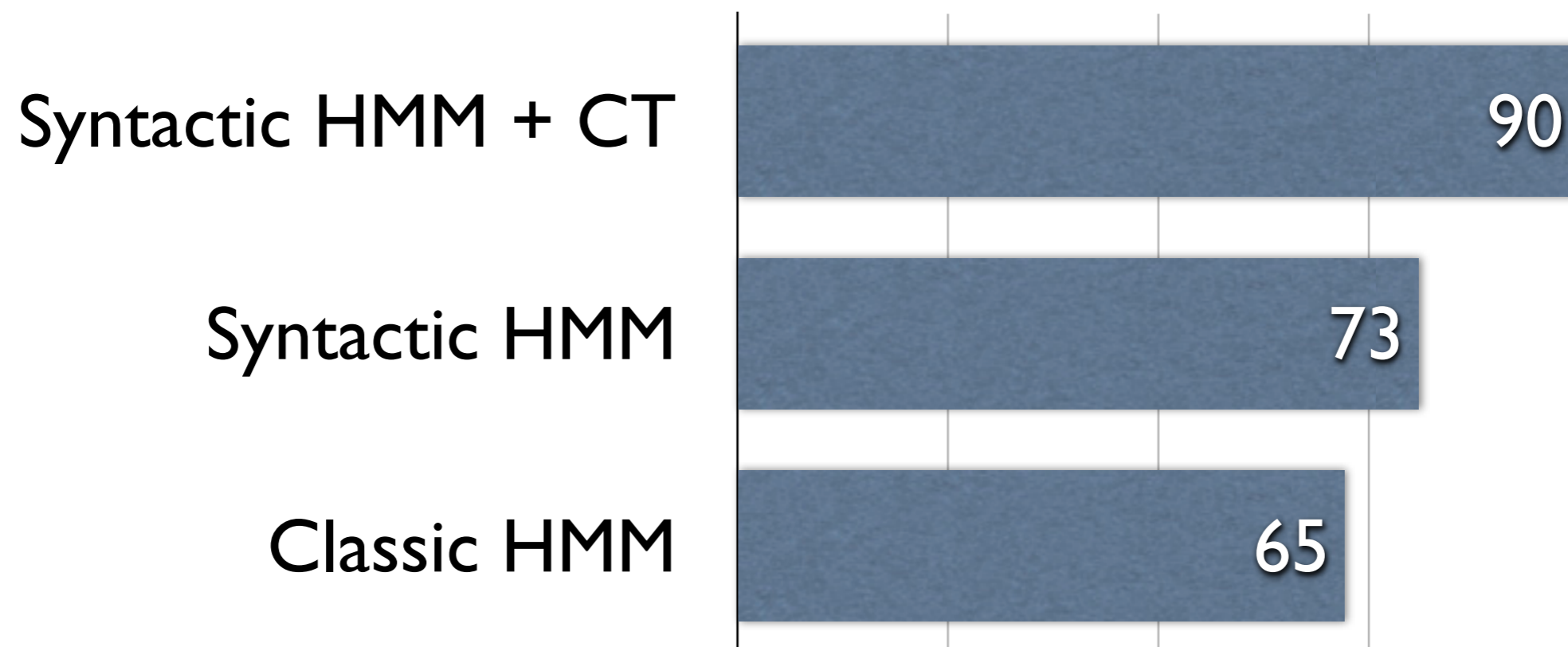
Classic HMM





Evaluation: Quantity of Rules Extracted

Rules extracted per sentence





Evaluation: Comparing Gold and Induced Rules

Evaluation Metric Idea:

Compare rules from gold alignments and induced alignments on both precision and recall.

Analog to the consistent phrase error rate (CPER) metric of Ayan & Dorr (06)



Evaluation: Comparing Gold and Induced Rules

Evaluation Metric Idea:

Compare rules from gold alignments and induced alignments on both precision and recall.

Analog to the consistent phrase error rate (CPER) metric of Ayan & Dorr (06)

	Precision	Recall	F1
Classic HMM	40.4	33.9	36.8
Syntactic HMM	41.3	36.7	38.9
Syntactic HMM + CT	39.6	41.1	40.3



Evaluation: Comparing Gold and Induced Rules

Evaluation Metric Idea:

Compare rules from gold alignments and induced alignments on both precision and recall.

Analog to the consistent phrase error rate (CPER) metric of Ayan & Dorr (06)

	Precision	Recall	F1
Classic HMM	40.4	33.9	36.8
Syntactic HMM	41.3	36.7	38.9
Syntactic HMM + CT	39.6	41.1	40.3

F1 Increase: 9.5% in Chinese; 18.7% in French

Summary

- Tree transducer extraction systems should be wary of constituent-violating alignment errors

Summary

- Tree transducer extraction systems should be wary of constituent-violating alignment errors
- Conditioning the HMM alignment model on a parse tree corrects some such errors

Summary

- Tree transducer extraction systems should be wary of constituent-violating alignment errors
- Conditioning the HMM alignment model on a parse tree corrects some such errors
- Decoding heuristics correct even more

Summary

- Tree transducer extraction systems should be wary of constituent-violating alignment errors
- Conditioning the HMM alignment model on a parse tree corrects some such errors
- Decoding heuristics correct even more
- The resulting rules are more faithful to the rule set that should be extracted

Summary

- Tree transducer extraction systems should be wary of constituent-violating alignment errors
- Conditioning the HMM alignment model on a parse tree corrects some such errors
- Decoding heuristics correct even more
- The resulting rules are more faithful to the rule set that should be extracted
- Future work: end-to-end translation (BLEU)



Coming 07/07/07: BerkeleyAligner Software Package

- Agreement training of IBM models, which reduces AER 32% relative to GIZA++ (Liang et al., 06)
- Syntactic distortion model (this paper)
- Posterior decoding heuristics (this paper)
- Evaluation code: searches for posterior thresholds, compares decoding methods, & tracks AER during training
- Easy integration with the Berkeley Parser
- Pure Java 1.5 will run on any platform

Check it out:

<http://nlp.cs.berkeley.edu/pages/WordAligner.html>

Thank You



denero@berkeley.edu