# Hierarchical Machine Translation

Miguel Rios

Universiteit van Amsterdam

April 27, 2018

# Content

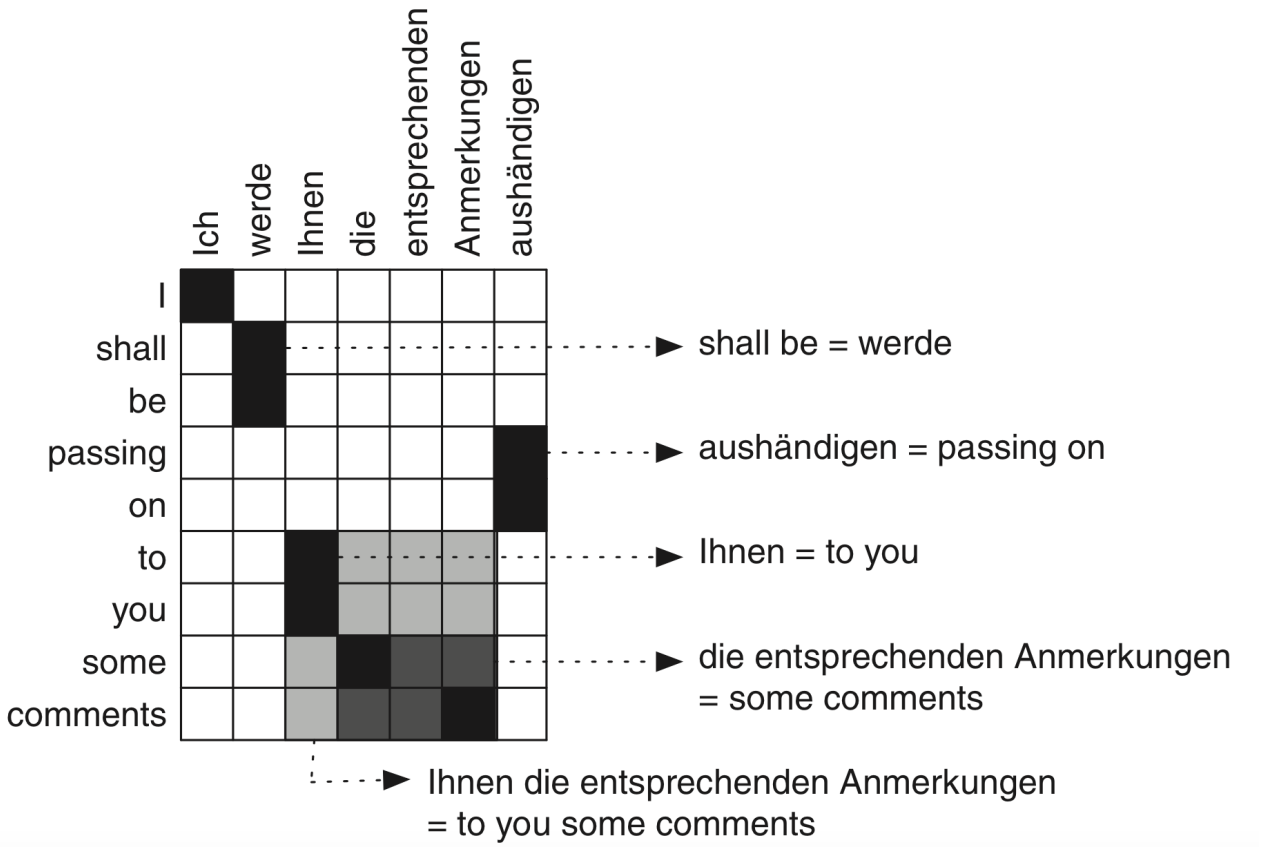



Figure: Koehn [2010]

werde $X$ aushändigen | shall be passing on $X$

# Why hierarchical structure?

Better generalisation

- compositionality
- reordering

# Why is reordering important?

Monotone translation is unrealistic

- languages differ wrt word-order

# Why is reordering important?

Monotone translation is unrealistic

- languages differ wrt word-order
  e.g. different syntactic structure

# Why is reordering important?

Monotone translation is unrealistic

- languages differ wrt word-order
  e.g. different syntactic structure
  e.g. rich morphology

# Why is reordering important?

Monotone translation is unrealistic

- languages differ wrt word-order
  e.g. different syntactic structure
  e.g. rich morphology

Reordering is arguably one of the hardest problems in MT

# Why is reordering important?

Monotone translation is unrealistic

- languages differ wrt word-order
  e.g. different syntactic structure
  e.g. rich morphology

Reordering is arguably one of the hardest problems in MT

- part of the model of translational equivalences
  *the part that determines the space of translations*

# Key aspects

Expressiveness

- how much can two languages differ wrt word order?

# Key aspects

Expressiveness

- how much can two languages differ wrt word order?

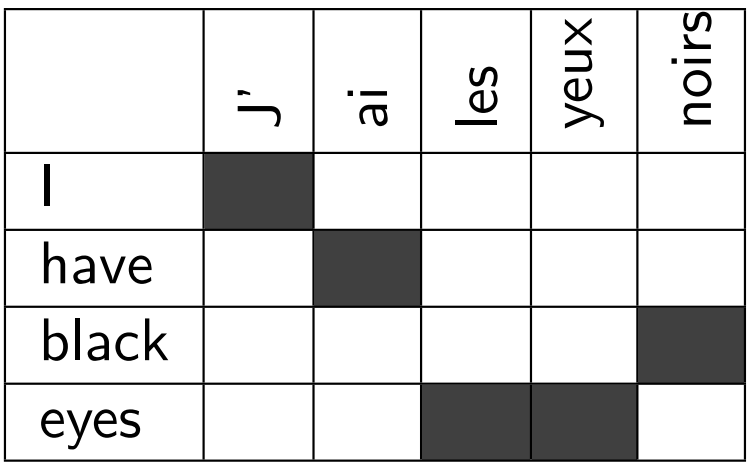
Modelling

- how many parameters do we have to estimate?

# Content

# Hierarchical phrase-based - Motivation

## Local Reordering



# Hierarchical phrase-based - Motivation

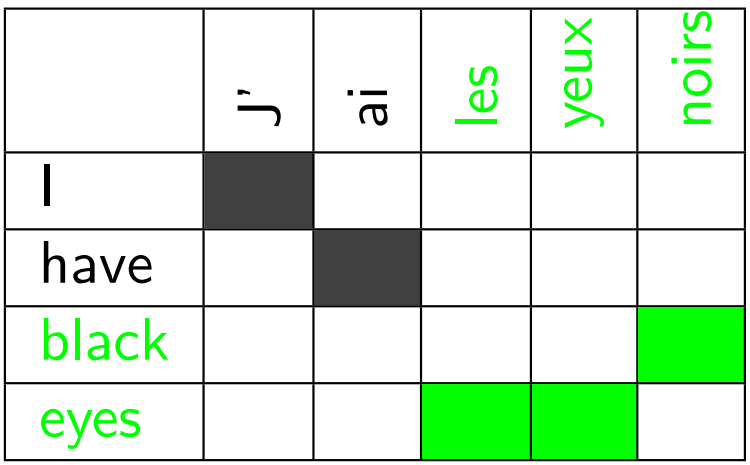
## Local Reordering

| | J' | ai | les | yeux | noirs |
|---|---|---|---|---|---|
| I | ■ | | | | |
| have | | ■ | | | |
| black | | | | | ■ |
| eyes | | | ■ | ■ | |

- Monotone
  J'$_1$ ai$_2$ $\rightarrow$ I$_1$ have$_2$

# Hierarchical phrase-based - Motivation

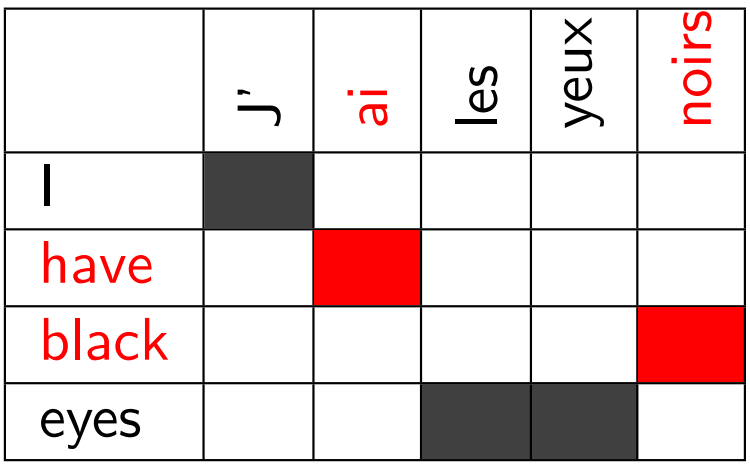
## Local Reordering




- Swap
  les yeux$_4$ noirs$_5$ $\rightarrow$ black$_3$ eyes$_4$

# Hierarchical phrase-based - Motivation

## Local Reordering




- Discontinuous
  ai$_2$ $X_{3-4}$ noirs$_5$ $\rightarrow$ have$_2$ black$_3$ $X_4$

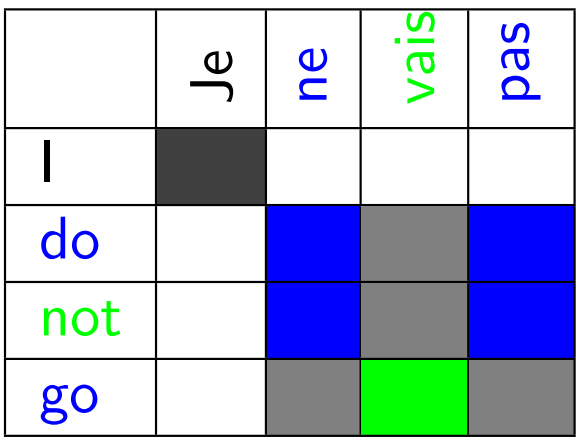
# Hierarchical phrase-based - Motivation

## Discontiguous Phrases

| | Je | ne | vais | pas |
|---|---|---|---|---|
| I | ■ | | | |
| do | | ■ | | ■ |
| not | | ■ | | ■ |
| go | | | ■ | |

# Hierarchical phrase-based - Motivation

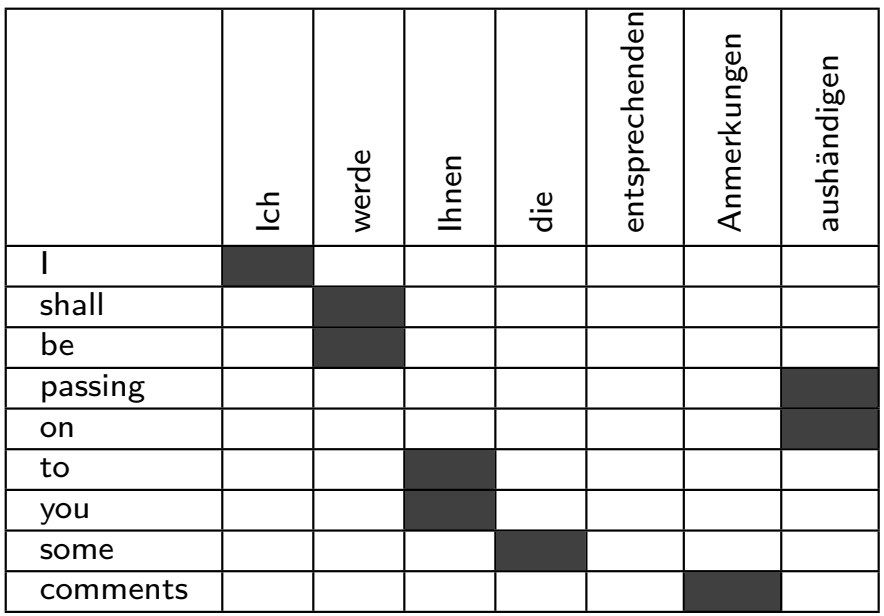
## Discontiguous Phrases




- Gappy phrase
  ne vais pas → do not go
  ne $X_{vais}$ pas → do not $X_{go}$

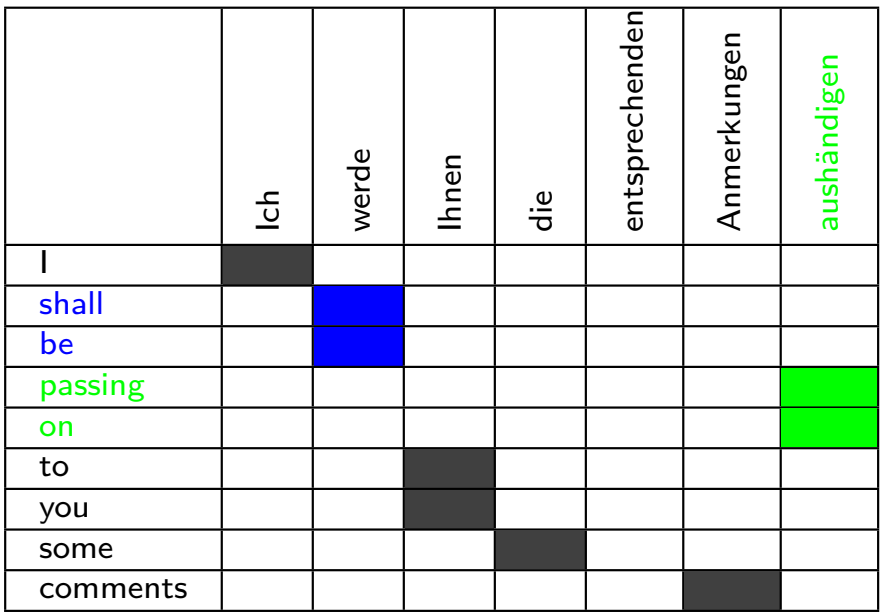
# Hierarchical phrase-based - Motivation

## Long Distance Reordering



# Hierarchical phrase-based - Motivation

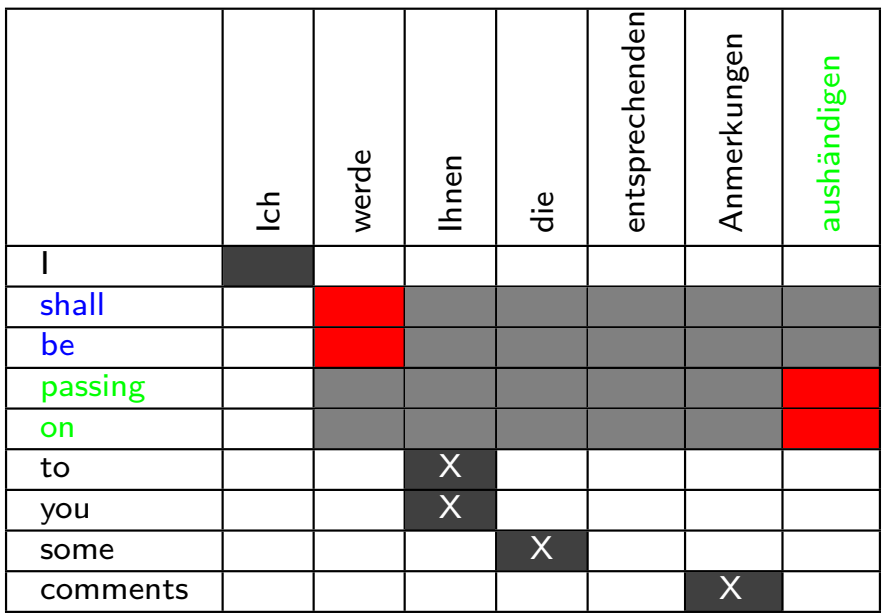
## Long Distance Reordering




- How can we extract a biphrase for **shall be passing on**?

# Hierarchical phrase-based - Motivation

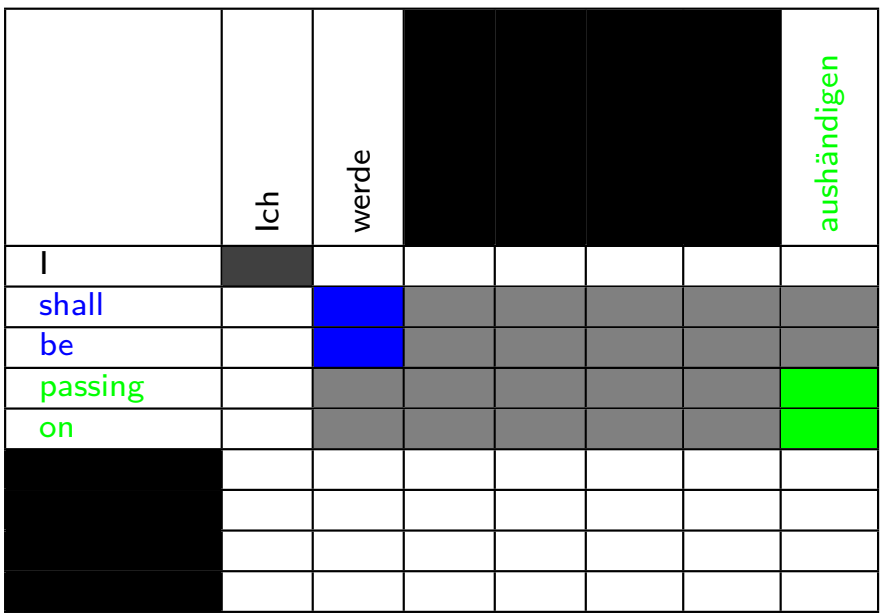
## Long Distance Reordering




- How can we extract a biphrase for **shall be passing on**?
- We cannot, we need to extract **to you some comments** along

# Hierarchical phrase-based - Motivation

## Long Distance Reordering




- How can we extract a biphrase for **shall be passing on**?
- We cannot, we need to extract **to you some comments** along
- Unless we replace all those words by a variable

# Hierarchical phrase-based - Motivation

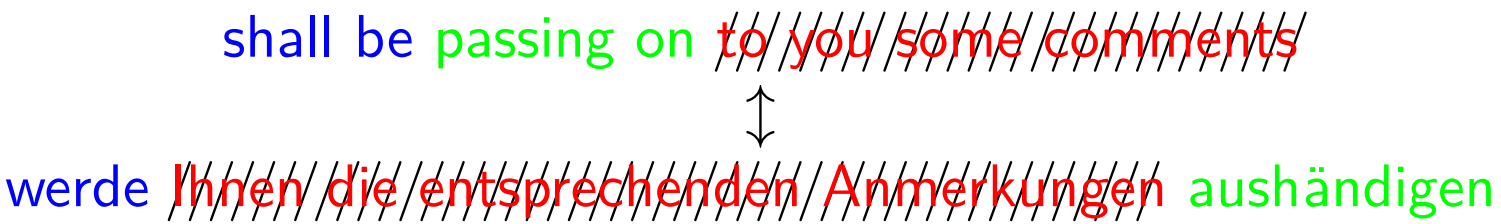
Long Distance Reordering

shall be passing on to you some comments

werde Ihnen die entsprechenden Anmerkungen aushändigen

# Hierarchical phrase-based - Motivation

Long Distance Reordering



# Hierarchical phrase-based - Motivation

Long Distance Reordering

shall be passing on $X$

$\updownarrow$

werde $X$ aushändigen

# Hiero

Extends phrase-based MT with hierarchical rules [Chiang, 2005]

# Hiero

Extends phrase-based MT with hierarchical rules [Chiang, 2005]

- conditions on word alignment

# Hiero

Extends phrase-based MT with hierarchical rules [Chiang, 2005]

- conditions on word alignment
- heuristic rule extraction

# Hiero

Extends phrase-based MT with hierarchical rules [Chiang, 2005]

- conditions on word alignment
- heuristic rule extraction
- heuristic scoring by relative frequency counting

# Hiero

Extends phrase-based MT with hierarchical rules [Chiang, 2005]

- conditions on word alignment
- heuristic rule extraction
- heuristic scoring by relative frequency counting
- log-linear model

# Hiero

Extends phrase-based MT with hierarchical rules [Chiang, 2005]

- conditions on word alignment
- heuristic rule extraction
- heuristic scoring by relative frequency counting
- log-linear model
- SCFG decoding

# Hiero

Extends phrase-based MT with hierarchical rules [Chiang, 2005]

- conditions on word alignment
- heuristic rule extraction
- heuristic scoring by relative frequency counting
- log-linear model
- SCFG decoding

# Hiero

Extends phrase-based MT with hierarchical rules [Chiang, 2005]

- conditions on word alignment
- heuristic rule extraction
- heuristic scoring by relative frequency counting
- log-linear model
- SCFG decoding

Motivation

- long-distance reordering

# Hiero

Extends phrase-based MT with hierarchical rules [Chiang, 2005]

- conditions on word alignment
- heuristic rule extraction
- heuristic scoring by relative frequency counting
- log-linear model
- SCFG decoding

Motivation

- long-distance reordering
- lexicalised reordering

# Heuristic rule extraction

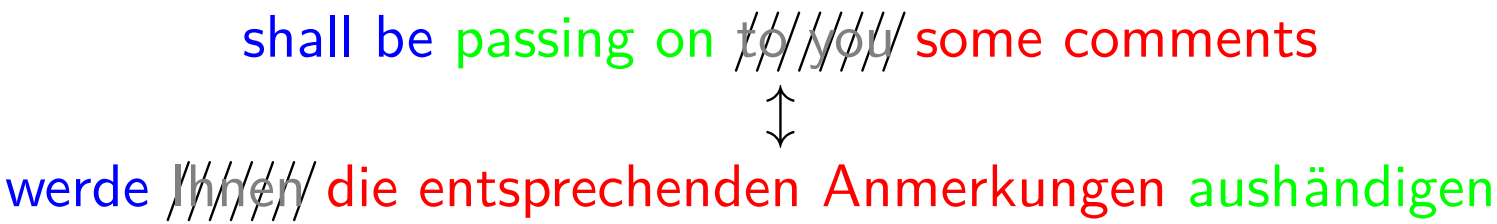
Initial phrase pairs created with same heuristic as PBSMT.

shall be passing on to you some comments

werde Ihnen die entsprechenden Anmerkungen aushändigen

# Heuristic rule extraction

Initial phrase pairs created with same heuristic as PBSMT.



# Heuristic rule extraction

Initial phrase pairs created with same heuristic as PBSMT.

shall be passing on $X_1$ some comments

werde $X_1$ die entsprechenden Anmerkungen aushändigen

# Heuristic rule extraction

Initial phrase pairs created with same heuristic as PBSMT.

shall be passing on $X_1$ some comments

↕

werde $X_1$ die entsprechenden Anmerkungen aushändigen

# Heuristic rule extraction

Initial phrase pairs created with same heuristic as PBSMT.

shall be passing on $X_1$ $X_2$

$\updownarrow$

werde $X_1$ $X_2$ aushändigen

# Heuristic rule extraction

Initial phrase pairs created with same heuristic as PBSMT.

$[X] \rightarrow$ shall be passing on $X_1$ $X_2$ | werde $X_1$ $X_2$ aushändigen

$[X] \rightarrow$ shall be passing on $X_3$ | werde $X_3$ aushändigen

$[X] \rightarrow$ to you | Ihnen

$[X] \rightarrow$ some comments | die entsprechenden Anmerkungen

$[X] \rightarrow$ to you some comments | Ihnen die entsprechenden Anmerkungen

# Hiero - Scoring

Relative frequency: assume all fragments have been "observed"
Give a count of one to phrase pair occurrence, then distribute its weight equally among the obtained rules.

- Joint rule probatility: $p(LHS, RHS_{source}, RHS_{target})$

$$p(X, \text{la maison } X_1, \text{the } X_1 \text{ house})$$

- Rule application probability: $p(RHS_{source}, RHS_{target}|LHS)$

$$p(\text{la maison } X_1, \text{the } X_1 \text{ house}|X)$$

- Direct translation probability: $p(RHS_{target}|RHS_{source}, LHS)$

$$p(\text{the } X_1 \text{ house}|\text{la maison } X_1, X)$$

- Noisy-channel translation probability: $p(RHS_{source}|RHS_{target}, LHS)$

$$p(\text{la maison } X_1|\text{the } X_1 \text{ house}, X)$$

- Lexical translation probability

$\prod_{t_i \in RHS_{target}} p(t_i|RHS_{source}, a)$ $\prod_{s_i \in RHS_{source}} p(s_i|RHS_{target}, a)$

# Hiero - Model

Log-linear combination of features

# Hiero - Model

Log-linear combination of features Linear model

$$S_\theta(e, d, f) = \theta^T \sum_{r_{s,t} \epsilon d} h_i(r_{s,t}|e, f)$$

where $s$ is a span over F,
$t$ is a span over E
and $r$ is a rule.
Weighted synchronous CFG.
LM.

# Hiero - Model

(0) J' (1) ai (2) les (3) yeux (4) noir (5)
(0) I (1) have (2) black (3) eyes (4)

$$X[0-2/0-2] \rightarrow \text{J' ai}|\text{I have}$$

$$X[2-4/3-4] \rightarrow \text{les yeux}|\text{eyes}$$

$$X[2-5/2-4] \rightarrow X[2-4]\text{noir}|\text{black}X[3-4]$$

$$S[0-5/0-4] \rightarrow X[0-2]X[2-5]|X[0-2]X[2-4]$$

# Content

# Decoding

Phrase-based

Tree-based

# Decoding

Phrase-based

- Left-to-Right

Tree-based

- Bottom-Up

# Decoding

Phrase-based

- Left-to-Right
- Beam Search

Tree-based

- Bottom-Up
- Chart Parsing (In the next Lab.)

# Decoding

Phrase-based

- Left-to-Right
- Beam Search
- Formally intersection:

Tree-based

- Bottom-Up
- Chart Parsing (In the next Lab.)
- Formally intersection:

# Decoding

Phrase-based

- Left-to-Right
- Beam Search
- Formally intersection:
- FST (TM) $\times$ FSA (LM)

Tree-based

- Bottom-Up
- Chart Parsing (In the next Lab.)
- Formally intersection:
- SCFG (TM) $\times$ FSA (LM)

# Questions?

# References I

David Chiang. A hierarchical phrase-based model for statistical machine translation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 263–270, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics. doi: 10.3115/1219840.1219873. URL `http://www.aclweb.org/anthology/P05-1033`.

Philipp Koehn. *Statistical Machine Translation*. Cambridge University Press, New York, NY, USA, 1st edition, 2010. ISBN 0521874157, 9780521874151.