

Hierarchical Machine Translation

Miguel Rios

Universiteit van Amsterdam

April 20, 2017

Content

- ① Motivation
- ② Hierarchical models of translation
Hiero
- ③ Decoding

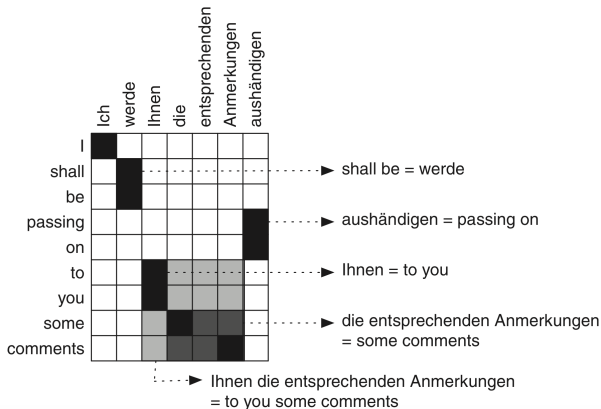


Figure: Koehn [2010]

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

- 1 Motivation
- 2 Hierarchical models of translation
Hiero
- 3 Decoding

Hierarchical phrase-based - Motivation

Local Reordering

	J'	ai	les	yeux	noirs
I					
have					
black					
eyes					

Hierarchical phrase-based - Motivation

Local Reordering

	J'	ai	les	yeux	noirs
I					
have					
black					
eyes					

- Monotone

$J'_1 \text{ ai}_2 \rightarrow I_1 \text{ have}_2$

Hierarchical phrase-based - Motivation

Local Reordering

	J'	ai	les	yeux	noirs
I					
have					
black					
eyes					

- Swap

les yeux₄ noirs₅ → black₃ eyes₄

Hierarchical phrase-based - Motivation

Local Reordering

	J'	ai	les	yeux	noirs
I					
have					
black					
eyes					

- Discontinuous

ai₂ X₃₋₄ noirs₅ → have₂ black₃
X₄

Hierarchical phrase-based - Motivation

Discontiguous Phrases

	Je	ne	vais	pas
I				
do				
not				
go				

Hierarchical phrase-based - Motivation

Discontiguous Phrases

	Je	ne	vais	pas
I				
do				
not				
go				

- Gappy phrase

ne vais pas → do not go

ne X_{vais} pas → do not X_{go}

Hierarchical phrase-based - Motivation

Long Distance Reordering

	Ich	werde	Ihnen	die	entsprechenden	Anmerkungen	aushändigen
I							
shall							
be							
passing							
on							
to							
you							
some							
comments							

Hierarchical phrase-based - Motivation

Long Distance Reordering

	Ich	werde	Ihnen	die	entsprechenden	Anmerkungen	aushändigen
I							
shall							
be							
passing							
on							
to							
you							
some							
comments							

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

Hierarchical phrase-based - Motivation

Long Distance Reordering

	Ich	werde	Ihnen	die	entsprechenden	Anmerkungen	aushändigen
I							
shall							
be							
passing							
on							
to			X				
you			X				
some				X			
comments						X	

- 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

	Ich	werde					aushändigen
I							
shall							
be							
passing							
on							

- 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

shall be passing on to you some comments
↕
werde Ihnen die entsprechenden Anmerkungen aushändigen

Hierarchical phrase-based - Motivation

Long Distance Reordering

shall be passing on *X*



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.

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.

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



werde X_1 X_2 aushändigen

Heuristic rule extraction

Initial phrase pairs created with same heuristic as PBSMT.

[X] → shall be passing on X_1 X_2 | werde X_1 X_2 aushändigen

[X] → shall be passing on X_3 | werde X_3 aushändigen

[X] → to you | Ihnen

[X] → some comments | die entsprechenden Anmerkungen

[X] → 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 probability: $p(LHS, RHS_{source}, RHS_{target})$

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

- 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

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

Hiero - Model

Log-linear combination of features

Hiero - Model

Log-linear combination of features Linear model

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

where s is a span over F

and t is a span over E

Weighted synchronous CFG.

LM.

Content

- 1 Motivation
- 2 Hierarchical models of translation
- 3 Decoding**

Decoding by Parsing

J' ai les yeux noirs

❶ PRP0/PRP \rightarrow J' | I

❷ JJ \rightarrow noirs | black

❸ NP0/NP \rightarrow ^{DT} les ^{NN} yeux JJ | JJ

❹ VP0/VP \rightarrow ^{VB} ai NP0 | ^{VB} have NP

❺ S \rightarrow PRP0 VP0 | PRP VP

Decoding by Parsing

J'_1 ai les yeux noirs

PRP0₁

|
 J'_1

PRP₁

|
 I_1

- ① PRP0/PRP $\rightarrow J' \mid I$
- ② JJ \rightarrow noirs \mid black
- ③ NP0/NP $\rightarrow \overset{DT}{les} \overset{NN}{yeux} JJ \mid JJ$
- ④ VP0/VP $\rightarrow \overset{VB}{ai} NP0 \mid \overset{VB}{have} NP$
- ⑤ S \rightarrow PRP0 VP0 \mid PRP VP

$\{I_1\}$

Decoding by Parsing

J'_1 ai les yeux noirs_2

PRP0₁

JJ₂

PRP₁

JJ₂

|

|

|

|

J'₁

noirs₂

I₁

black₂

① PRP0/PRP → J' | I

② JJ → noirs | black

③ NP0/NP → $\overset{DT}{les} \overset{NN}{yeux}$ JJ | JJ

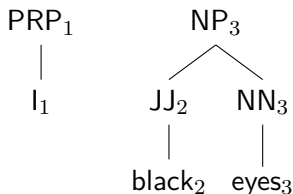
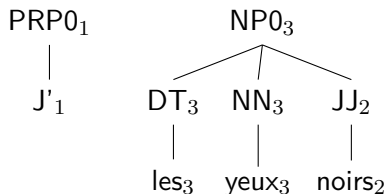
④ VP0/VP → $\overset{VB}{ai}$ NP0 | $\overset{VB}{have}$ NP

⑤ S → PRP0 VP0 | PRP VP

{I₁, black₂}

Decoding by Parsing

J'_1 ai les $yeux_3$ $noirs_2$

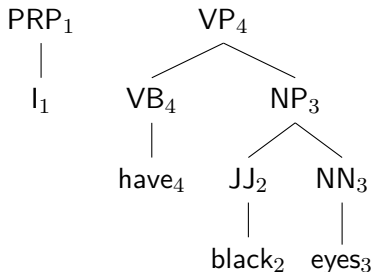
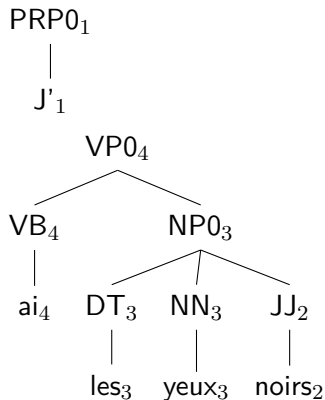


- 1 PRP0/PRP → J' | I
- 2 JJ → noirs | black
- 3 NP0/NP → $\overset{DT}{les}$ $\overset{NN}{yeux}$ JJ | JJ
- 4 VP0/VP → $\overset{VB}{ai}$ NP0 | $\overset{VB}{have}$ NP
- 5 S → PRP0 VP0 | PRP VP

{ I_1 , $black_2$ $eyes_3$ }

Decoding by Parsing

J'₁ ai₄ les yeux₃ noirs₂

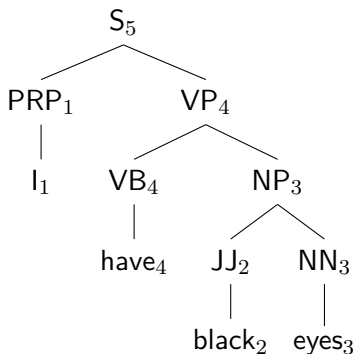
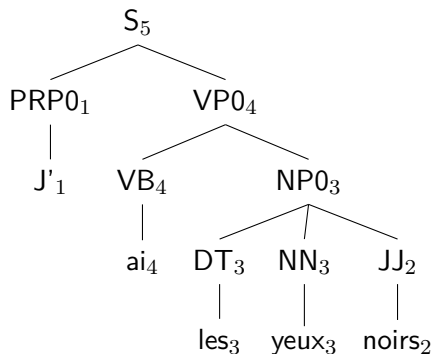


- ① PRP0/PRP → J' | I
- ② JJ → noirs | black
- ③ NP0/NP → ^{DT}les ^{NN}yeux JJ | JJ
- ④ VP0/VP → ^{VB}ai NP0 | ^{VB}have NP

{I₁, have₄ black₂ eyes₃}

Decoding by Parsing

J'_1 ai_4 les_3 $yeux_3$ $noirs_2$



- ① PRP0/PRP → J' | I
- ② JJ → noirs | black
- ③ NP0/NP → ^{DT}les ^{NN}yeux JJ | JJ
- ④ VP0/VP → ^{VB}ai NP0 | ^{VB}have NP
- ⑤ S → PRP0 VP0 | PRP VP

{ I_1 $have_4$ $black_2$ $eyes_3$ }

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:
- $\text{FST (TM)} \times \text{FSA (LM)}$

Tree-based

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

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.