Natural Language Processing

Class 11:
An Introduction to Machine Translation

25 May 2018

Katrien Beuls / Paul Van Eecke

VRIJE UNIVERSITEIT BRUSSEL

Overview

What is Machine Translation (MT) and **what** is it used for?

Why is MT difficult?

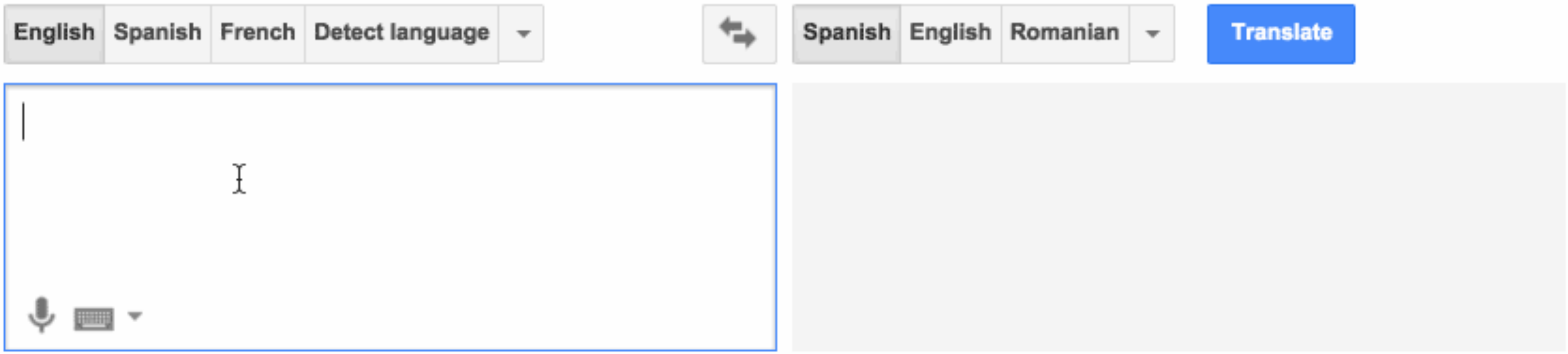
How is Machine Translation **achieved**?

- Rule-based
- Statistical
- Neural

How are translation results **evaluated**?

- Bleu, NIST

What is Machine Translation (MT)



https://medium.com/@ageitgey/machine-learning-is-fun-part-5-language-translation-with-deep-learning-and-the-magic-of-sequences-2ace0acca0aa

What can MT be used for?

NOT for translation tasks that require:

- a deep and rich understanding of the source language input
- a deep and rich understanding of the text/domain/world
- a creative command of the target language (e.g. literature)

is samen met

24 april om 18:36 ·

Onze chip 😍❤️😍

· Origineel weergeven · Beoordeel deze vertaling

is samen met

24 april om 18:36 ·

Notre puce 😍❤️😍

Onze chip 😍❤️😍

· Origineel verbergen · Beoordeel deze vertaling

What can MT be used for?

SOMETIMES for non-literary translation tasks, in particular when

- only a rough translation is needed (gist)
  (e.g. Most uses of Google Translate)

- there is still a human post-editor (CAT)
  (speeding up de translation process)

- it is limited to a sub-language domain in which fully automatic high-quality translation (FAHQT) is achievable
  (famous example: Canadian French-English weather reports)

Why is MT difficult?

Languages are very different on many levels

-> lexically

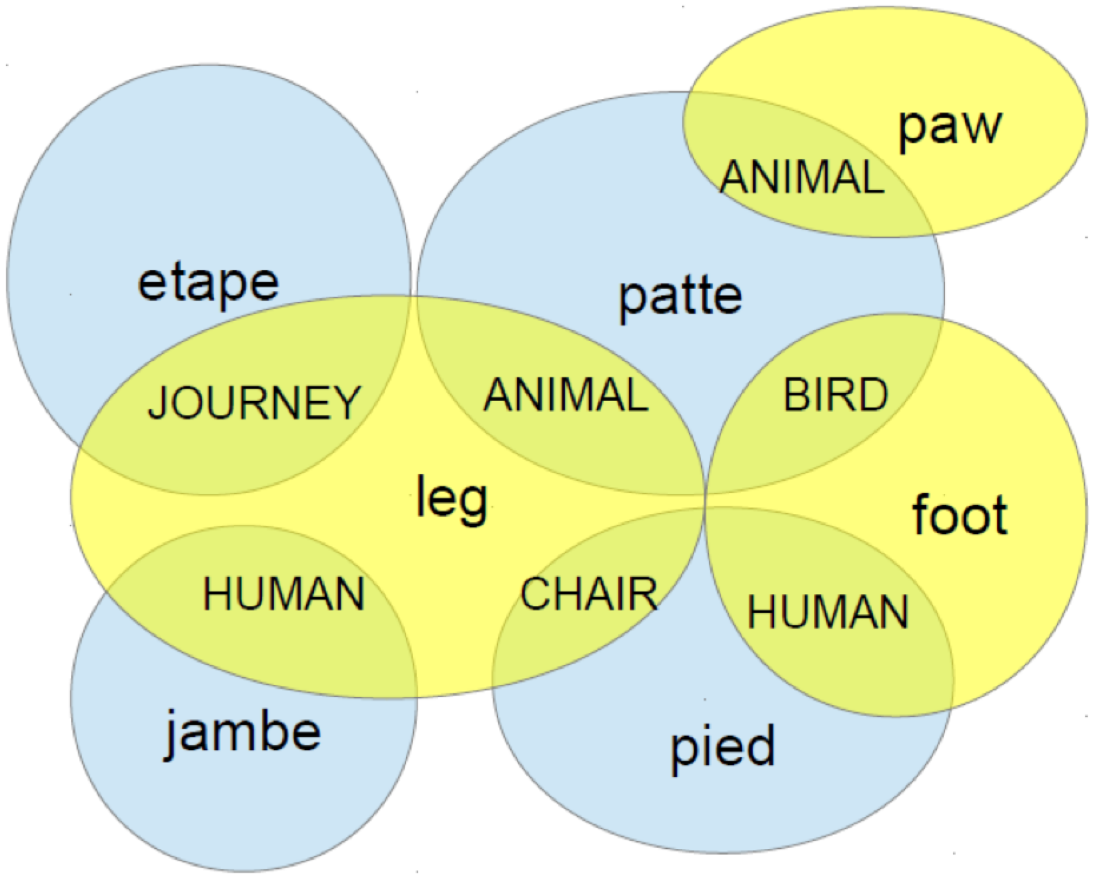
-> Morphologically

-> Syntactically

-> Semantically

Why is MT difficult?

Lexically:



Why is MT difficult?

Morphologically:

English: the man’s house

Hungarian: the man house-his

Why is MT difficult?

Syntactically:

English: The bottle floated out

Spanish: La botellia salio flotando
(the bottle exited while floating)

Why is MT difficult?

Semantically:

German: Die Lehrerin hat mich das gefragt.
English: The teacher asked me that.

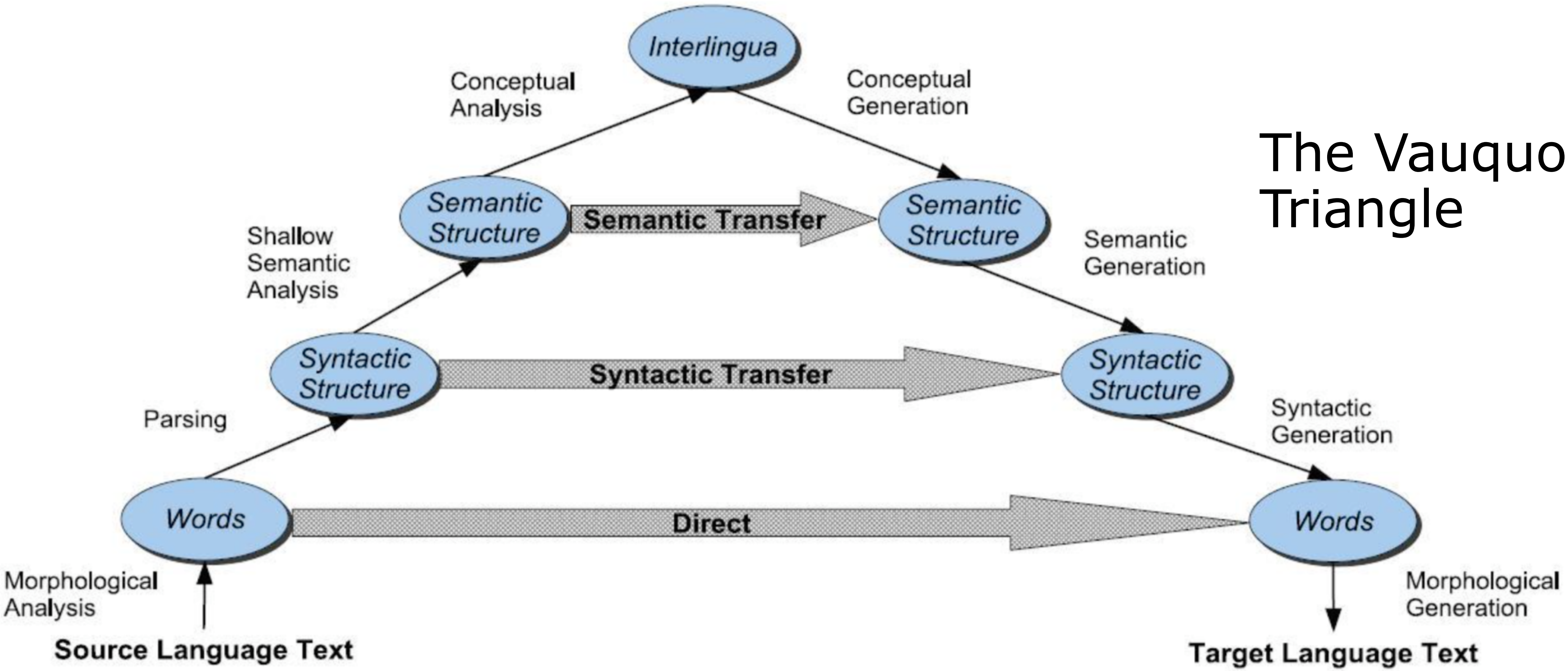
(The female teacher?)
(What if we translate from English into german?)

French: Elle a gaspillé son argent
English: She wasted his?/her? money.

How is MT Achieved?

- Rule-Based
- Statistical
- Neural

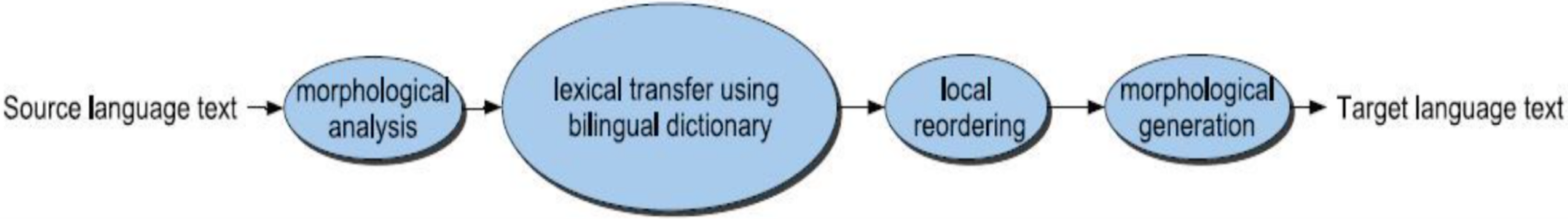
Rule-Based Machine Translation



The Vauquois Triangle

1. Direct Translation

- Word per word, with a few enhancements:
- Only very shallow analyses



- Not really usable:
  - Needs very complex translation dictionary
    - (e.g. ‘slap’ -> ‘dar una bofetada a’)
  - lack of long-distance reorderings

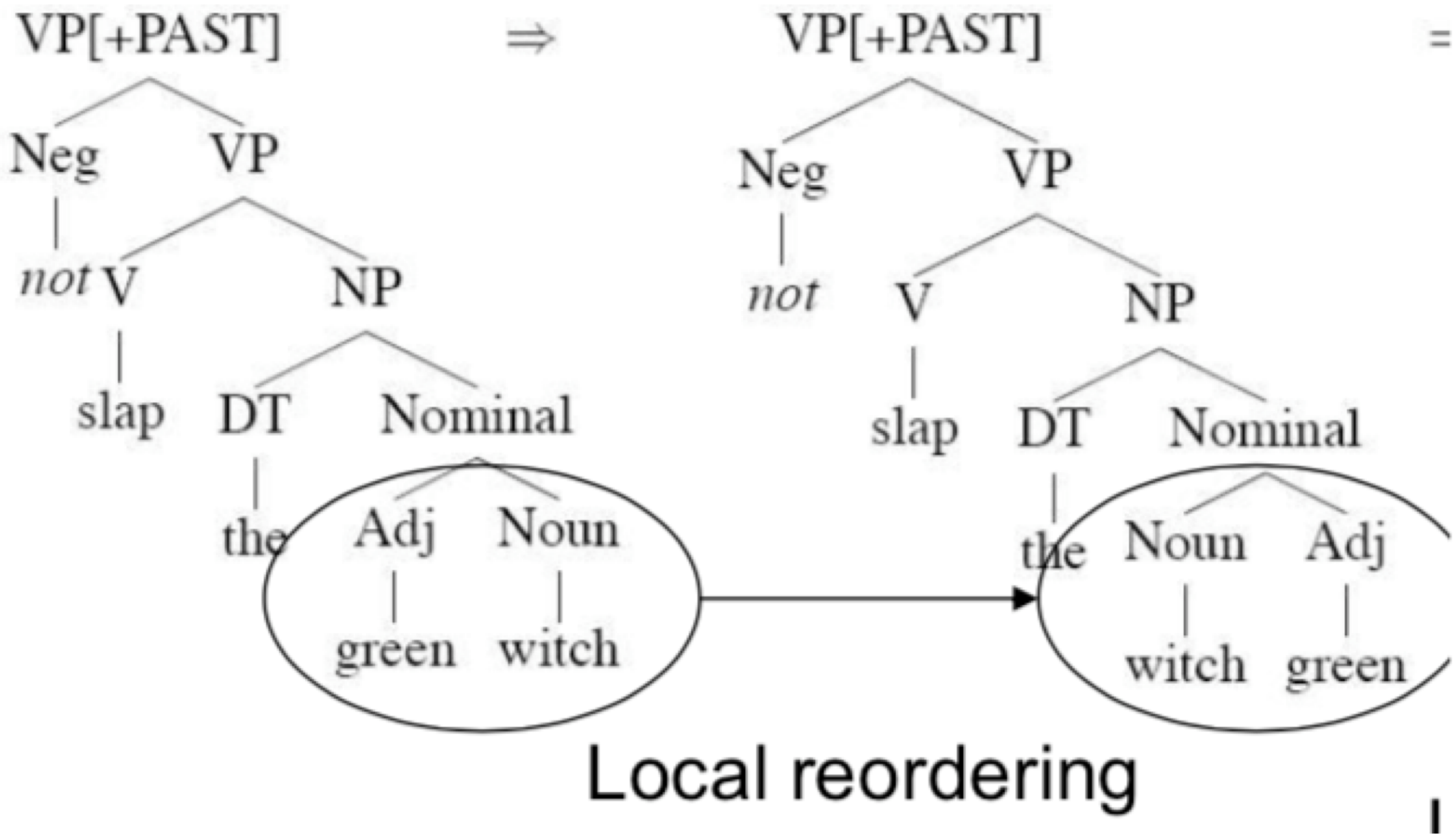
2. Transfer-Based Translation

Using knowledge of the differences between two languages

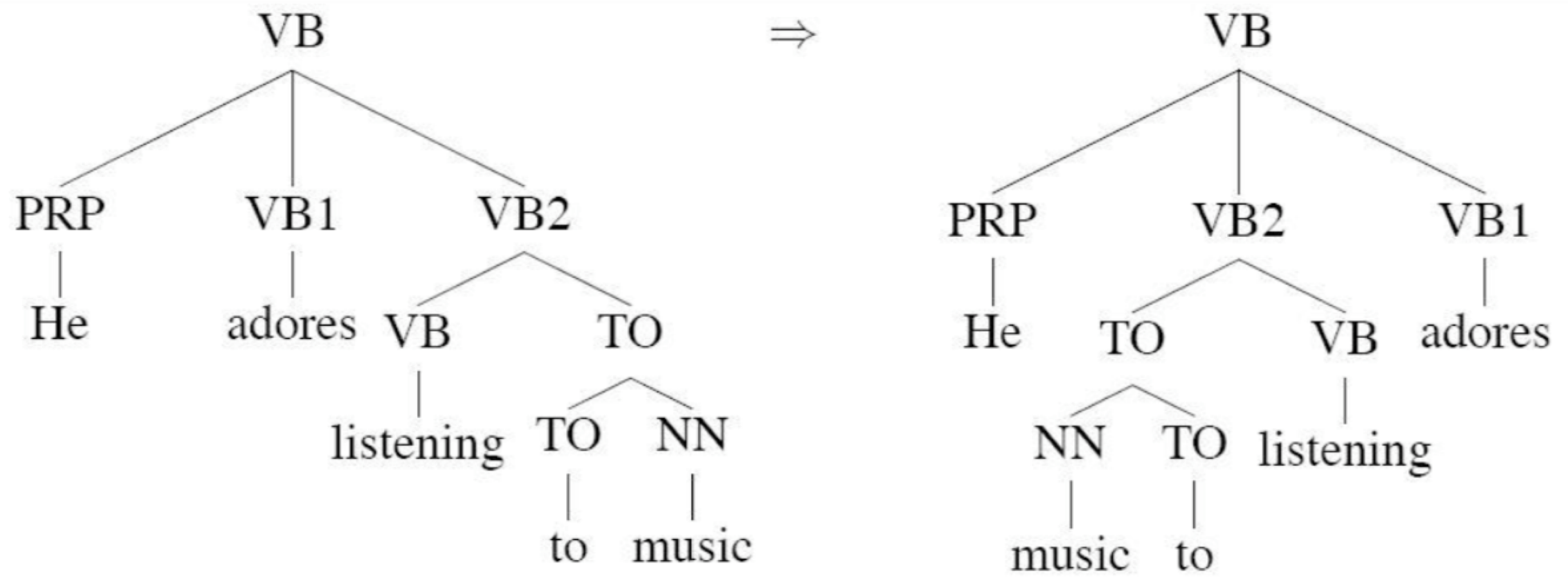
- Analysis of input (syntax tree, some semantics)
- Transfer

- Generation of output (syntax tree, some semantics)

2. Transfer-Based Translation



2. Transfer-Based Translation



3. Interlingua

- Using a grammar for the source language: mapping of a sentence to an abstract, language-independent meaning representation
(= **comprehension**).

- Using a grammar for the target language: mapping of a meaning representation to an utterance
(= **formulation**).

3. Interlingua

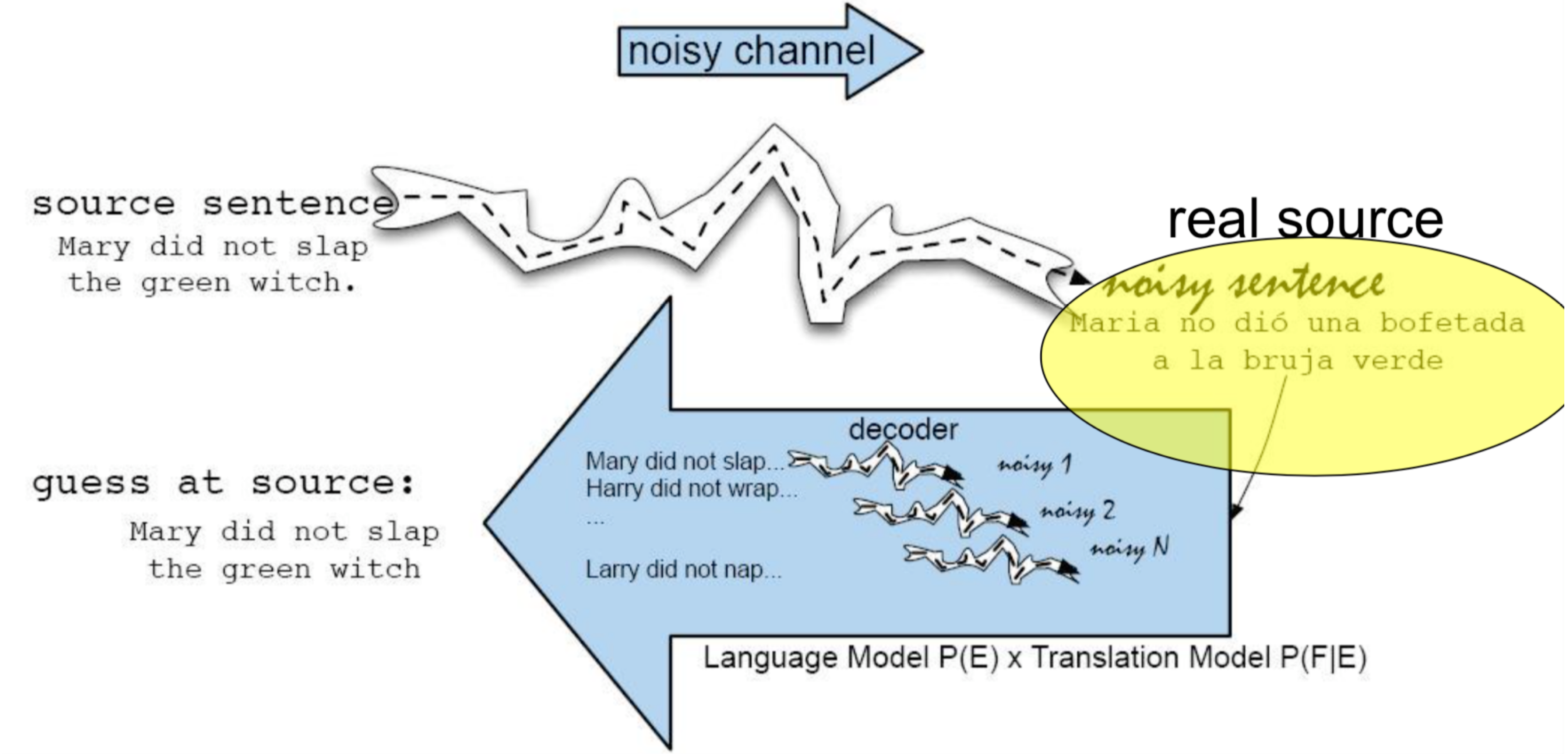
$$
\begin{bmatrix}
\text{EVENT} & \text{SLAPPING} \\
\text{AGENT} & \text{MARY} \\
\text{TENSE} & \text{PAST} \\
\text{POLARITY} & \text{NEGATIVE} \\
\text{THEME} & \begin{bmatrix}
\text{WITCH} & \\
\text{DEFINITENESS} & \text{DEF} \\
\text{ATTRIBUTES} & \begin{bmatrix} \text{HAS-COLOR} & \text{GREEN} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

3. Interlingua

- **Tools:** e.g. bidirectional grammar formalisms such as **Fluid Construction Grammar**

- **Problems**:
    - what is the interlingua? How to construct a meaning representation underlying all languages?
    - Exact semantic parsing is a problem far from solved

Statistical Machine Translation



Statistical Machine Translation

E: Utterance in English
F: Utterance in a Foreign Language

$$\hat{E} = argmax_E \, P(E|F)$$

$$\hat{E} = argmax_E \, \frac{P(F|E)\,P(E)}{P(F)}$$

$$\hat{E} = argmax_E \, \underbrace{P(F|E)}_{\text{translation model}} \, \underbrace{P(E)}_{\text{language model}}$$

Statistical Machine Translation

- Translation model = combination of
  - Translation probability
  - Distortion probability (for other word orders)

Statistical Machine Translation

Position	1	2	3	4	5
English	Mary	did not	slap	the	green witch
Spanish	Maria	no	dió una bofetada	a la	bruja verde

- Distortion are all 1
- P(F|E) = P(Maria,Mary) x d(1) x

  P(no|did not) x d(1) x

  P(dió una bofeatada|slap) x d(1) x

  P(a la|the) x d(1) x

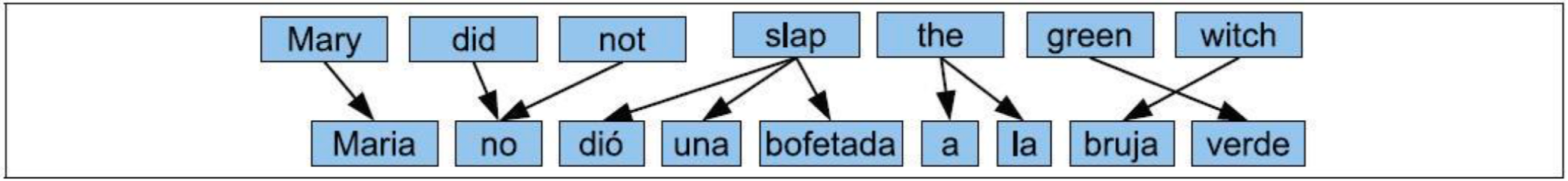
  P(bruja verde|green witch) x d(1)

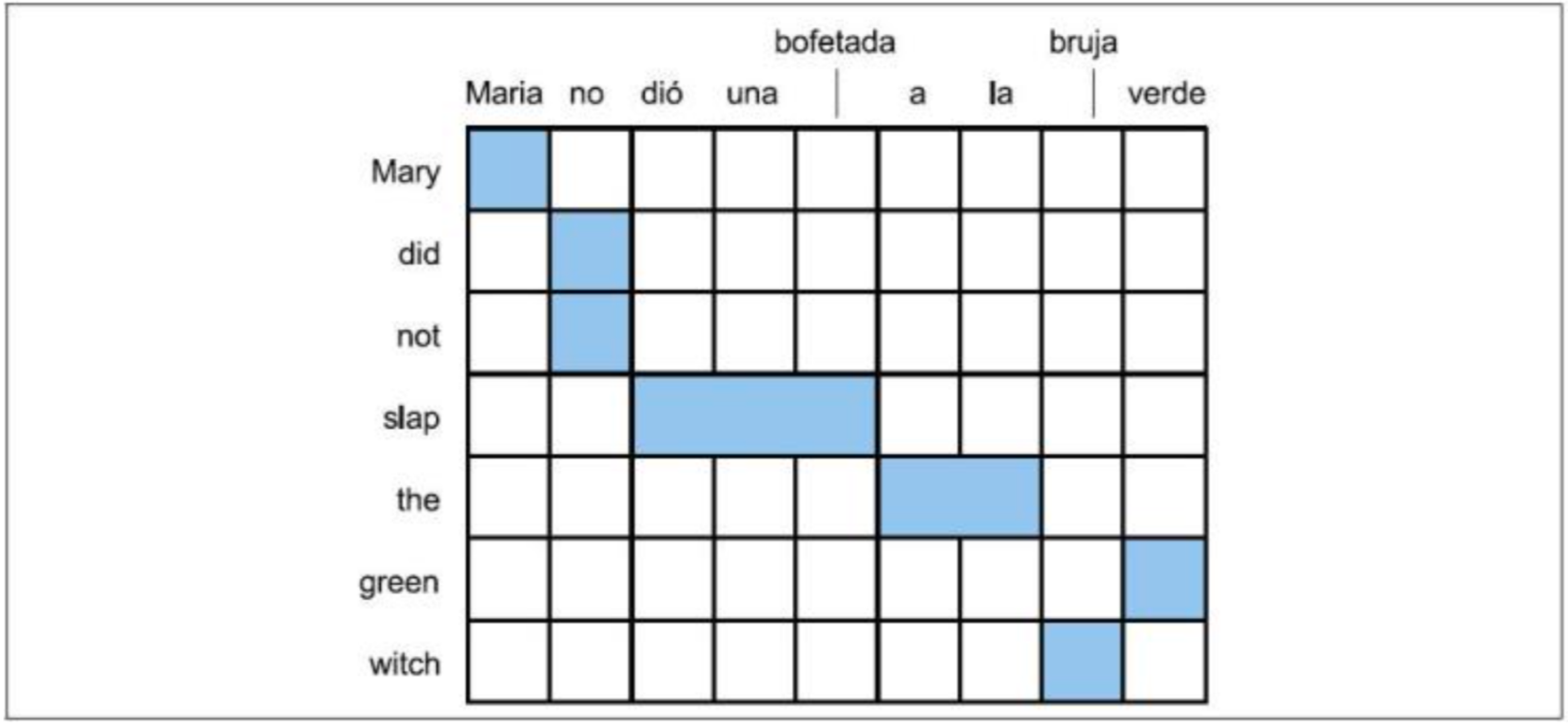
Statistical Machine Translation

- Problem: There exist no large-scale parallel corpora with phrases annotated.

- Solution: Compute phrase alignments based on word alignments that are trained from sentence-aligned parallel corpora

Word Alignment Models





IBM Model 1

- Expectation-Maximization
- Imagine this parallel corpus:

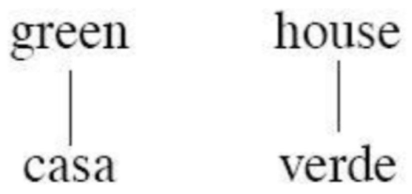
  green house – casa verde

  the house - la casa
- E = {green, house , the} ; S = {casa, verde ,la}
- Start with Uniform probabilities

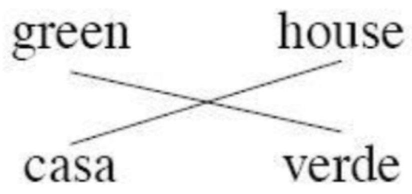
$t(\text{casa}\mid\text{green}) = \frac{1}{3}$	$t(\text{verde}\mid\text{green}) = \frac{1}{3}$	$t(\text{la}\mid\text{green}) = \frac{1}{3}$
$t(\text{casa}\mid\text{house}) = \frac{1}{3}$	$t(\text{verde}\mid\text{house}) = \frac{1}{3}$	$t(\text{la}\mid\text{house}) = \frac{1}{3}$
$t(\text{casa}\mid\text{the}) = \frac{1}{3}$	$t(\text{verde}\mid\text{the}) = \frac{1}{3}$	$t(\text{la}\mid\text{the}) = \frac{1}{3}$

IBM Model 1

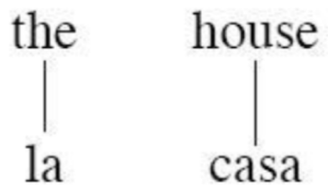
- Expectation Step 1a
- Compute P(a,f|e), multiplying all t probabilities



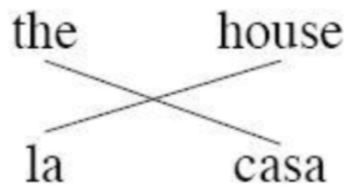
$$P(a,f|e) = t(\text{casa,green}) \times \text{t(verde,house)} = \frac{1}{3} \times \frac{1}{3} = \frac{\mathbf{1}}{\mathbf{9}}$$



$$P(a,f|e) = t(\text{verde,green}) \times \text{t(casa,house)} = \frac{1}{3} \times \frac{1}{3} = \frac{\mathbf{1}}{\mathbf{9}}$$



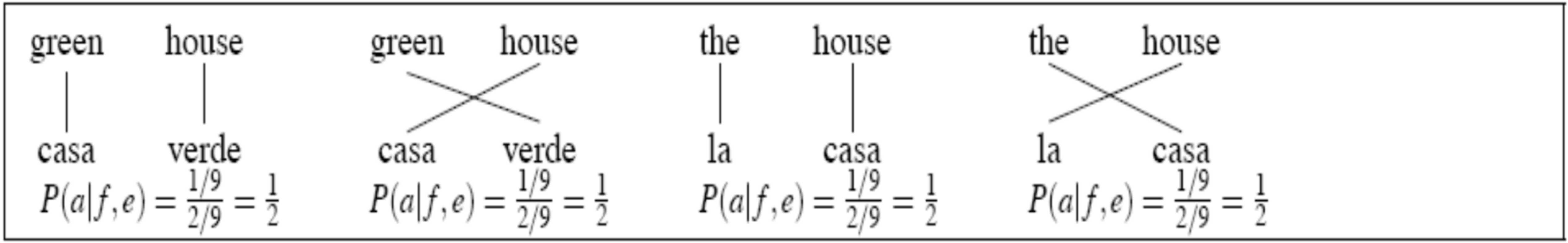
$$P(a,f|e) = t(\text{la,the}) \times \text{t(casa,house)} = \frac{1}{3} \times \frac{1}{3} = \frac{\mathbf{1}}{\mathbf{9}}$$



$$P(a,f|e) = t(\text{casa,the}) \times \text{t(la,house)} = \frac{1}{3} \times \frac{1}{3} = \frac{\mathbf{1}}{\mathbf{9}}$$

IBM Model 1

- Expectation Step 1b
- Normalize P(a,f|e)



IBM Model 1

- Expectation Step 1c
- Compute expected counts

$\text{tcount}(\text{casa}\|\text{green}) = \frac{1}{2}$	$\text{tcount}(\text{verde}\|\text{green}) = \frac{1}{2}$	$\text{tcount}(\text{la}\|\text{green}) = 0$	$\text{total}(\text{green}) = 1$
$\text{tcount}(\text{casa}\|\text{house}) = \frac{1}{2} + \frac{1}{2}$	$\text{tcount}(\text{verde}\|\text{house}) = \frac{1}{2}$	$\text{tcount}(\text{la}\|\text{house}) = \frac{1}{2}$	$\text{total}(\text{house}) = 2$
$\text{tcount}(\text{casa}\|\text{the}) = \frac{1}{2}$	$\text{tcount}(\text{verde}\|\text{the}) = 0$	$\text{tcount}(\text{la}\|\text{the}) = \frac{1}{2}$	$\text{total}(\text{the}) = 1$

IBM Model 1

- Maximization Step 1
- Compute MLE probabilities by normalizing tcounts to sum to 1

$t(\text{casa}\mid\text{green}) = \frac{1/2}{1} = \frac{1}{2}$	$t(\text{verde}\mid\text{green}) = \frac{1/2}{1} = \frac{1}{2}$	$t(\text{la}\mid\text{green}) = \frac{0}{1} = 0$
$t(\text{casa}\mid\text{house}) = \frac{1}{2} = \frac{1}{2}$	$t(\text{verde}\mid\text{house}) = \frac{1/2}{2} = \frac{1}{4}$	$t(\text{la}\mid\text{house}) = \frac{1/2}{2} = \frac{1}{4}$
$t(\text{casa}\mid\text{the}) = \frac{1/2}{1} = \frac{1}{2}$	$t(\text{verde}\mid\text{the}) = \frac{0}{1} = 0$	$t(\text{la}\mid\text{the}) = \frac{1/2}{1} = \frac{1}{2}$

IBM Model 1

- Expectation Step 2a
- Compute P(a,f|e), multiplying all t probabilities

green house	green house	the house	the house
casa verde	casa verde	la casa	la casa
$P(a,f\|e) = t(\text{casa,green}) \times t(\text{verde,house}) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$	$P(a,f\|e) = t(\text{verde,green}) \times t(\text{casa,house}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$	$P(a,f\|e) = t(\text{la,the}) \times t(\text{casa,house}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$	$P(a,f\|e) = t(\text{casa,the}) \times t(\text{la,house}) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$

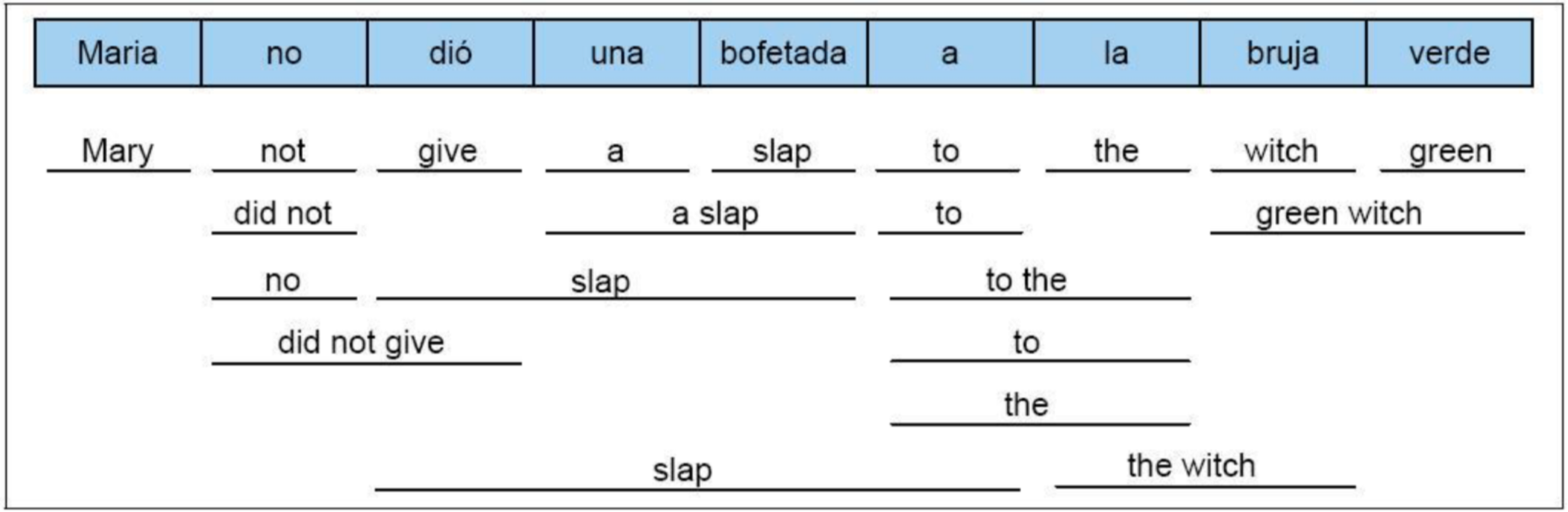
Statistical Machine Translation

- From word alignments, we can compute phrase alignments: **symmetrizing**

- **Decoding:** Given a translation model P(F|E) and a language model P(E), we want the highest E

  $$\hat{E} = \text{argmax}_e P(F|E)\, P(E)$$

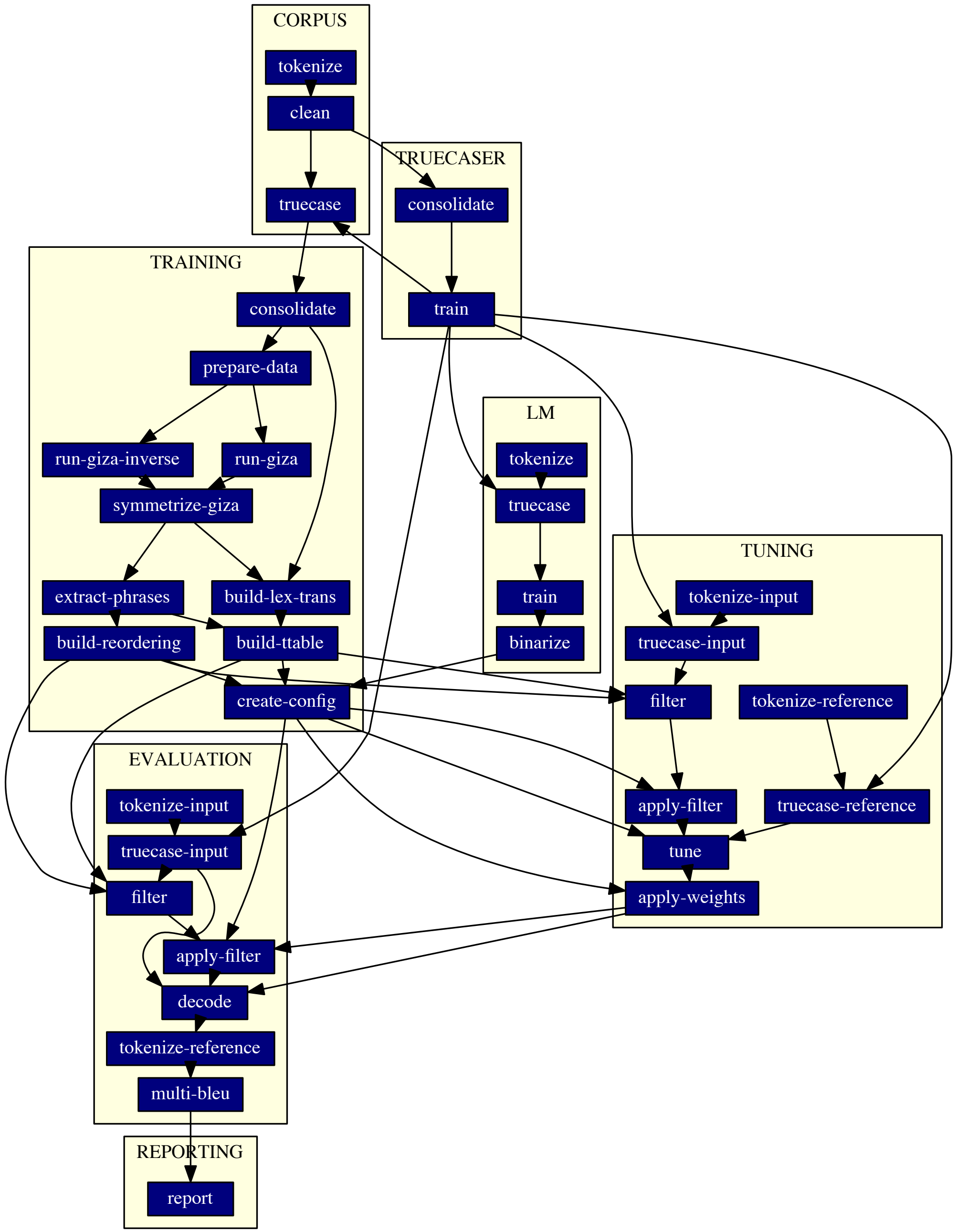
- This is a search process (e.g. A*)

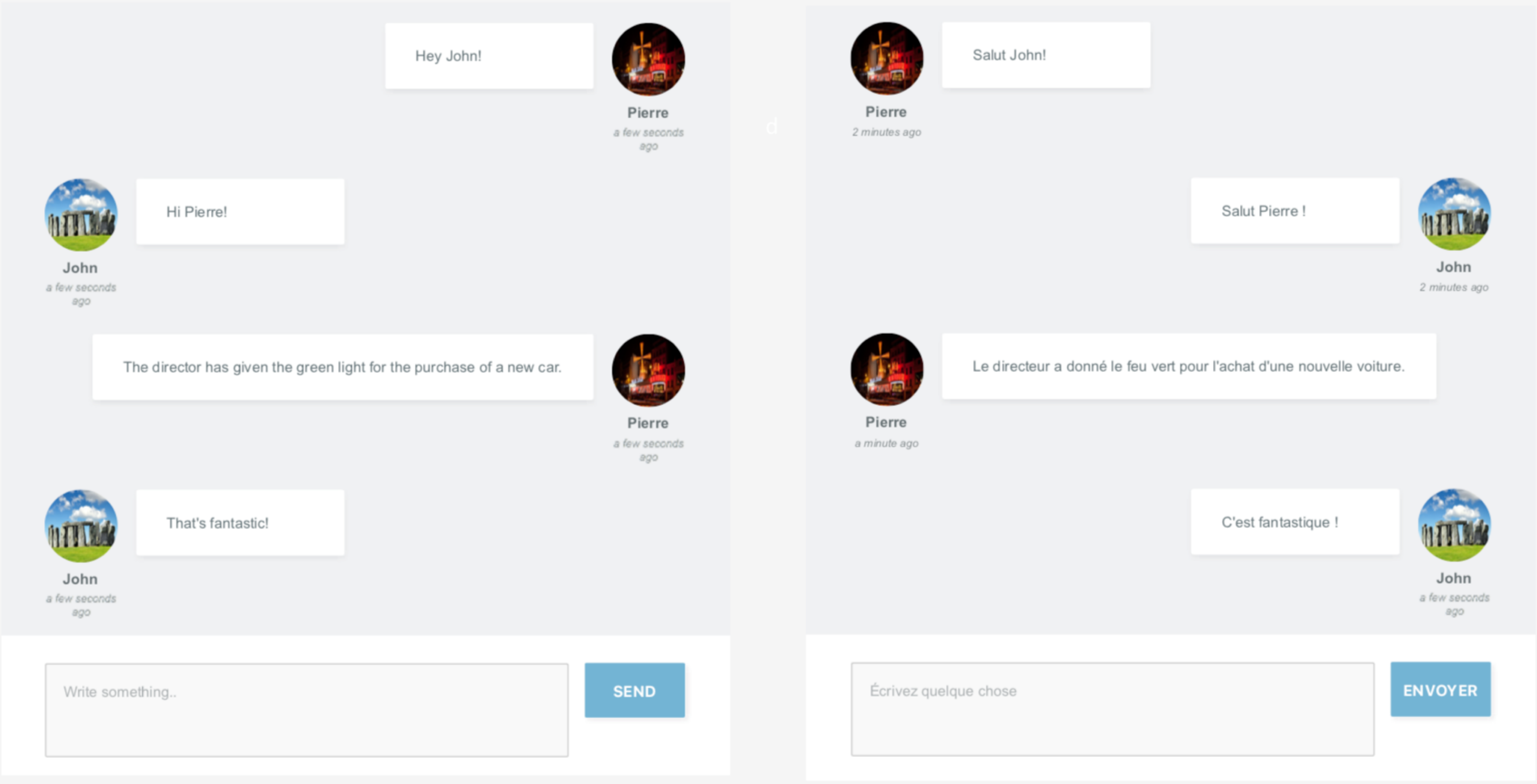

Maria
no
dió
una
bofetada
a
la
bruja
verde
Mary
not
give
a
slap
to
the
witch
green
did not
a slap
to
green witch
no
slap
to the
did not give
to
the
slap
the witch

SMT: Try it yourself

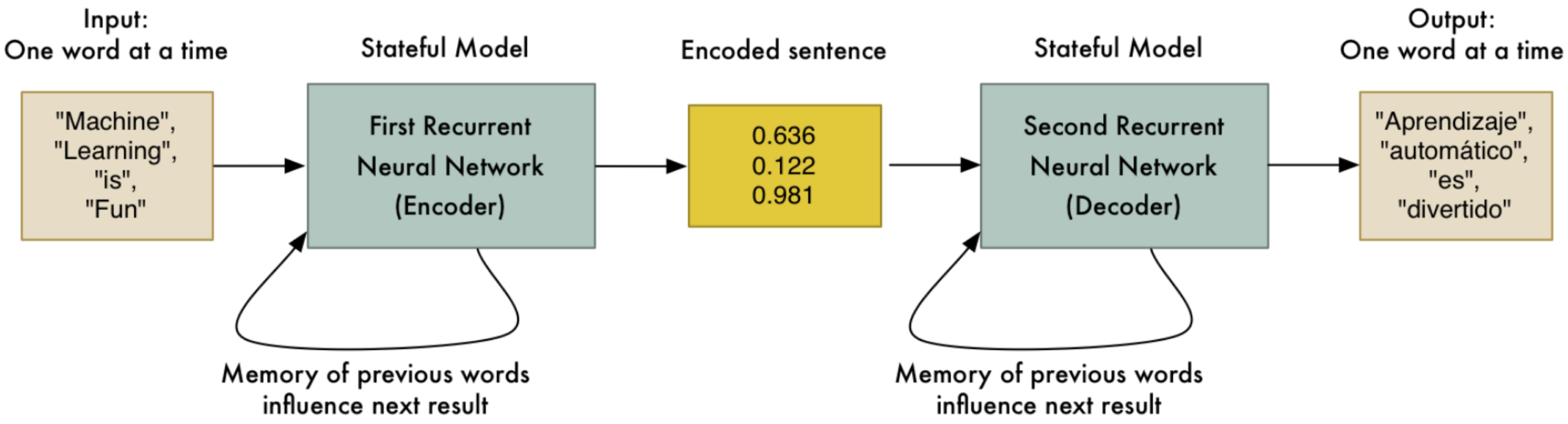
- Open source machine translation system
http://www.statmt.org/moses/

- Download parallel corpora (e.g. movie subtitles, European parliament, …)
- Preprocessing (perl scripts for tokenizing, truecasing, cleaning)
- Making a language model (SRILM, KENLM, …)
- Making phrase tables (giza++)
- Decoding (Moses decoder)


CORPUS
tokenize
clean
truecase
TRUECASER
consolidate
train
TRAINING
consolidate
prepare-data
run-giza-inverse
run-giza
symmetrize-giza
extract-phrases
build-lex-trans
build-reordering
build-ttable
create-config
LM
tokenize
truecase
train
binarize
TUNING
tokenize-input
truecase-input
filter
tokenize-reference
apply-filter
truecase-reference
tune
apply-weights
EVALUATION
tokenize-input
truecase-input
filter
apply-filter
decode
tokenize-reference
multi-bleu
REPORTING
report


Hey John!
Pierre
a few seconds ago
Hi Pierre!
John
a few seconds ago
The director has given the green light for the purchase of a new car.
Pierre
a few seconds ago
That's fantastic!
John
a few seconds ago
Write something..
SEND
Salut John!
Pierre
2 minutes ago
Salut Pierre !
John
2 minutes ago
Le directeur a donné le feu vert pour l'achat d'une nouvelle voiture.
Pierre
a minute ago
C'est fantastique !
John
a few seconds ago
Écrivez quelque chose
ENVOYER

Neural Machine Translation



NMT: Try it yourself

- Open source neural machine translation system
- http://opennmt.net

- LuaTorch, PyTorch or Tensorflow

Evaluating Machine Translation

- Human Raters (time-consuming and expensive!)
    - Fluency
    - Clarity
    - Naturalness
    - Style

    - Fidelity
    - Adequacy
    - Informativeness

    - Post-editing cost

Evaluating Machine Translation

- Automatic metrics (much worse, but some correlation with human judgements)

- **BLEU**, NIST, TER,METEOR

Evaluating Machine Translation

- BLEU (BiLingual Evaluation Understudy) is a precision metric: it is computed based on the number of words that a candidate translation shares with human references.

Cand 1:	It is a guide to action which ensures that the military always obeys the commands of the party
Cand 2:	It is to insure the troops forever hearing the activity guidebook that party direct
Ref 1:	It is a guide to action that ensures that the military will forever heed Party commands
Ref 2:	It is the guiding principle which guarantees the military forces always being under the command of the Party
Ref 3:	It is the practical guide for the army always to heed the directions of the party

Evaluating Machine Translation

- But simple precision does not work (7/7).
- Clip precision by maximum in reference count (2/7).

Candidate:	the	the	the	the	the	the	the
Reference 1:	the	cat	is	on	the	mat	
Reference 2:	there	is	a	cat	on	the	mat

- Everyone knows BLEU has its problems (but has been optimizing his/her models for years on BLEU)