Approaches to Machine Translation: Rule-based, Statistical and Hybrid

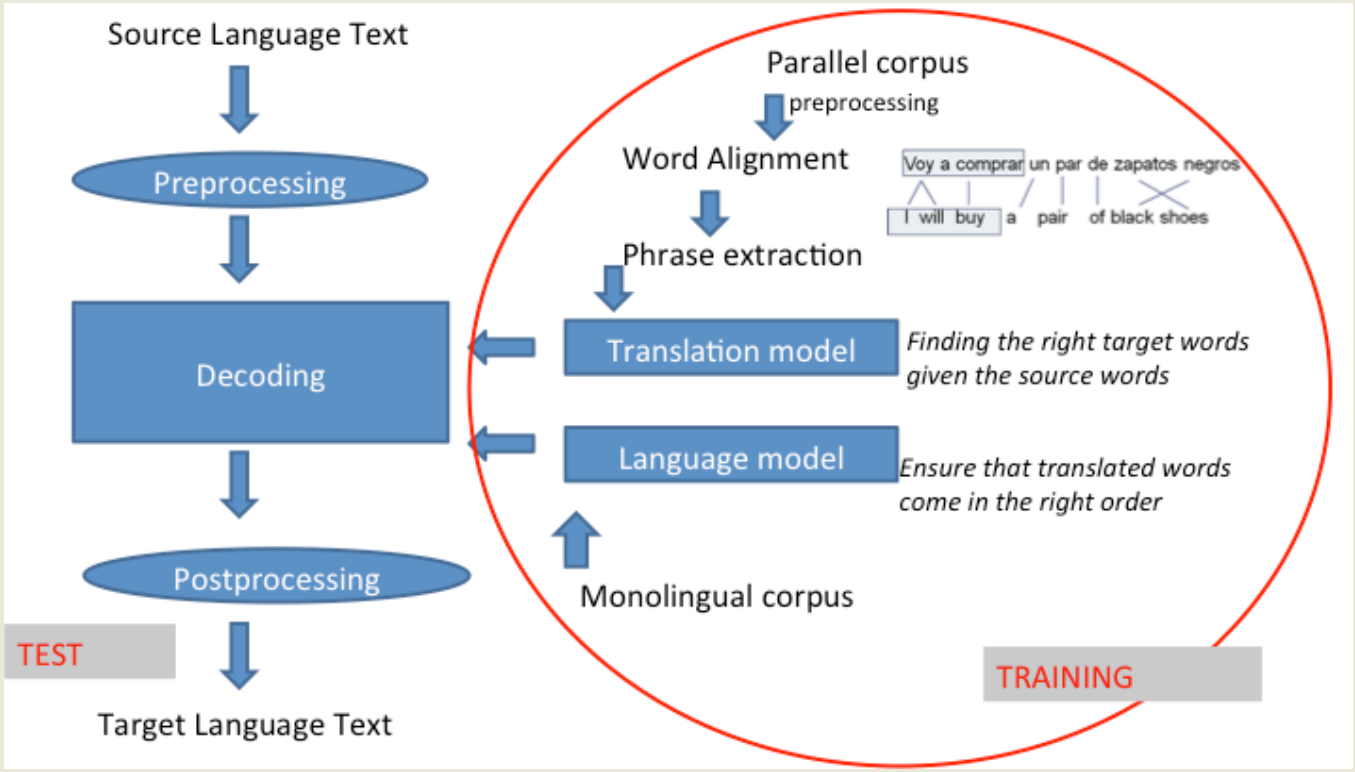
Alignment - Introduction (I)

Lluís Formiga

July 11, 2014

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

A picture is worth a million equations



Concepts to be studied

- Noisy Channel Model
- Lexical translation
- Word Alignment
- Expectation Maximization (EM) Algorithm
- IBM Models 1--5
  - IBM Model 1: lexical translation
  - IBM Model 2: alignment model
  - IBM Model 3: fertility
  - IBM Model 4: relative alignment model
  - IBM Model 5: deficiency
- HMM Models: dependent alignment model
- Problems of Word Alignment
- Quality of Word Alignment

From the Noisy Channel model we have:

$$p(e|f) =$$

From the Noisy Channel model we have:

$$p(e|f) = \arg\max_{e} \mathbf{p(f|e)}p(e)$$

From the Noisy Channel model we have:

$$p(e|f) = \arg\max_e \mathbf{p(f|e)}p(e)$$

However...

- We don't have a model for $p(f|e)$
- Does this model depend on previous decisions?

Word based models

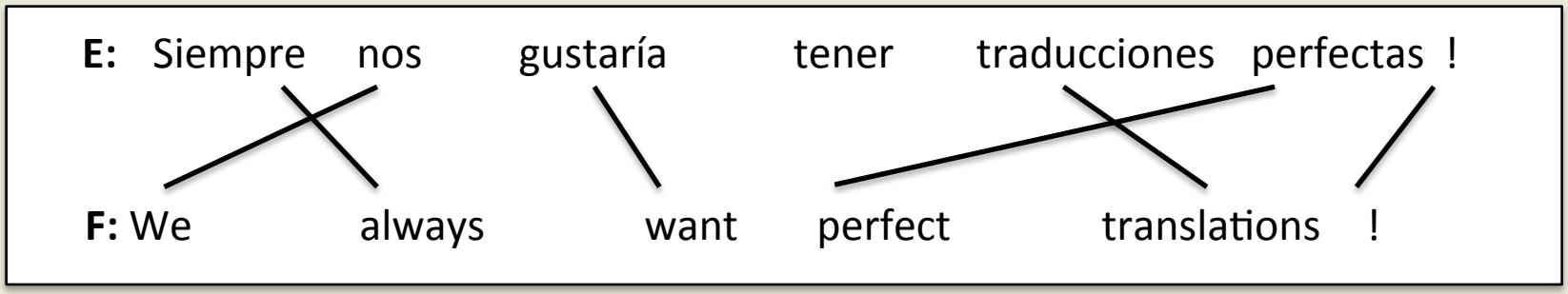
We can collect basic statistics:

- IBM Model 1 is just a table capturing $t(f|e)$

Translations of *mesa*	$p(f|e)$
table	0.3771
round	0.1476
panel	0.1344
round-table	0.0452
petitioners	0.0282
bureau	0.0229
officers	0.0190
Committee	0.0169
Round	0.0153
roundtable	0.0124

What is alignment? Alignment function

- When translating, we align words between languages.
- What is alignment?
  - Each foreign language word f is generated by exactly by one translated language word e

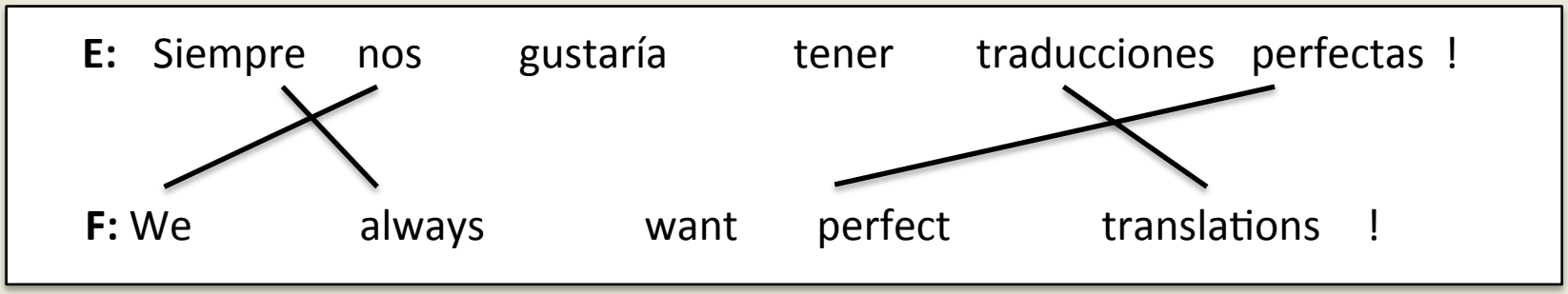


- We need to define an alignment function
- Not an easy task! Different phenomena might occur!

Concepts to deal with Problems:Reordering

- The order of the source does not match with the target

Former example:

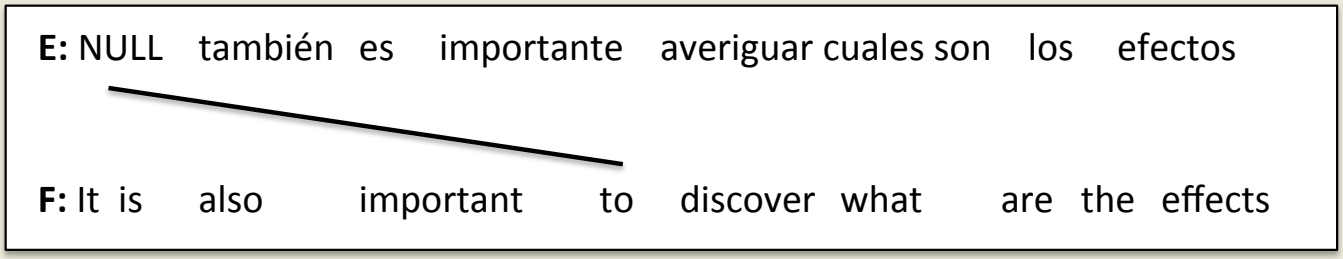


$$\mathsf{a} = \langle 2, 1, \cdot, 6, 5, \cdot \rangle$$

Concepts to deal with Problems: Word insertion

- Words (typically function words) are added when translating
- Special NULL token helps map translated words to source

New example:

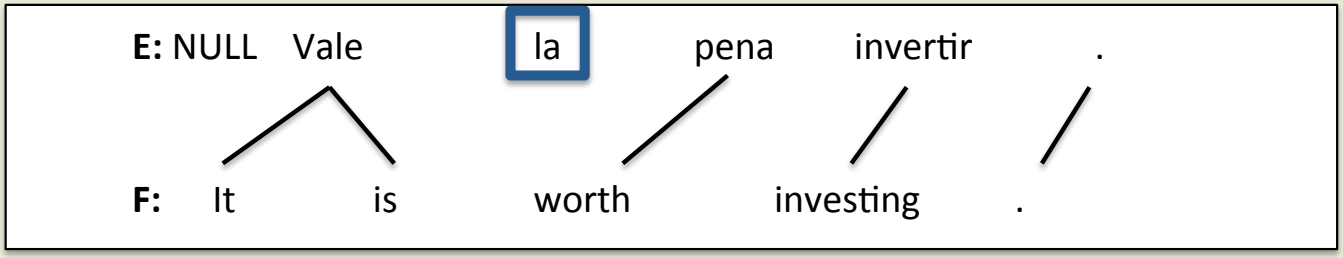


$$\mathbf{a} = \langle \cdot, \cdot, \cdot, \cdot, 0, \cdot, \cdot, \cdot, \cdot, \cdot \rangle$$

Concepts to deal with Problems: Word deletion

- Words may be dropped when translated
  (la disappears on English Sentence)

Former example:

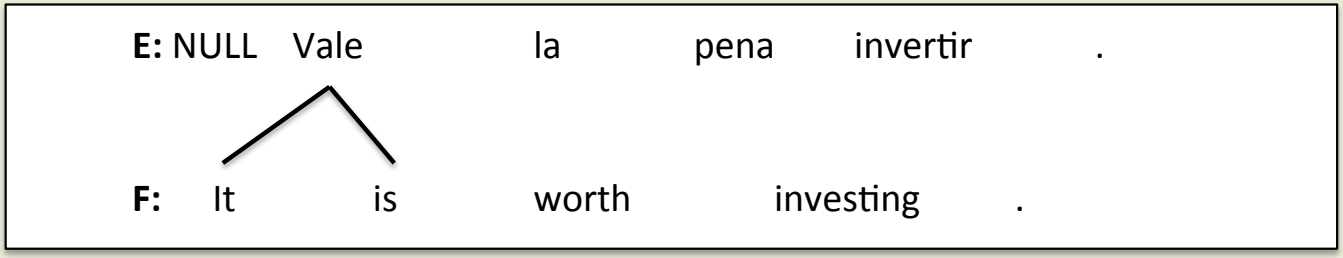


$$\mathbf{a} = \langle 1, 1, 3, 4, 5 \rangle$$

Concepts to deal with Problems: Fertility

- One-to-many translation
- A source word may translate into multiple target words:

Former example:



$$\mathbf{a} = \langle 1, 1, \cdot, \cdot, \cdot \rangle$$

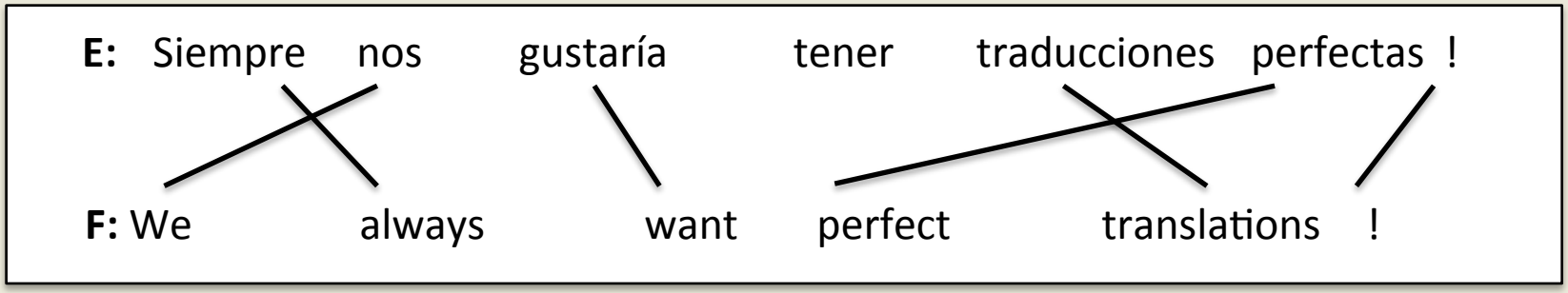
What is alignment? Alignment Function

- Maps each foreign word f at position j

  $\rightarrow$ to translated word e at position i with function $a : j \rightarrow i$

foreign: source (e.g. English)
translated: target (e.g. Spanish)

Example:



$$a = \langle 2, 1, 3, 6, 5, 7 \rangle$$
$$l_f = 6 \quad \text{source token length}$$
$$l_e = 7 \quad \text{target token length}$$

What is alignment?

Basic statistics:

- IBM Model 1 captures $p(f|e)$

Translations of *mesa*	$p(f\|e)$
table	0.3771
round	0.1476
panel	0.1344
round-table	0.0452
petitioners	0.0282
bureau	0.0229
officers	0.0190
Committee	0.0169
Round	0.0153
roundtable	0.0124

Also we can collect advanced statistics:

- IBM Model 2 captures $q(j|i, l_f, l_e)$

j	i	l_f	l_e	$q(j\|i, l_e, l_f)$
1	1	5	7	0.27
1	2	5	7	0.14
⋮	⋮	5	7	0.07
5	7	5	7	1e-14
1	1	5	8	0.32
1	2	5	8	0.18
⋮	⋮	5	6	0.13
5	8	5	8	1e-19
⋮	⋮	⋮	⋮	⋮
1	1	6	8	0.30
1	2	6	8	0.12
⋮	⋮	6	8	0.17
6	8	6	8	1e-10

What is alignment?

Concept of alignment

- Alignments are obtained by means of unsupervised learning (Expectation Maximization Algorithm)
- Not a unique solution:
  - Each foreign word f has $l_e + 1$ choices, so we have l_f^{l+1} total combinations.
  - Hence, we built a conditional model projecting translations from the alignments
- We assume independence:
  - Every word is translated independently:

$$p(f_1, f_2, \ldots, f_{l_f} | e_1, e_2, \ldots, e_{l_e}, l_f) =$$
$$= \sum_{a \in A} p(f_1, \ldots, f_{l_f}, a_1, \ldots, a_{l_f} | e_1, \ldots, e_{l_e} l, l_f)$$

IBM Alignment Models

- Proposed by IBM in the late 80s/early 90s
- Five different models:

IBM Model 1	lexical translation (words);
IBM Model 2	adds absolute align. model
IBM Model 3	adds fertility model;
IBM Model 4	adds relative align. model
IBM Model 5	fixes deficiency

Next session

- ~~Noisy Channel Model~~
- ~~Lexical translation~~
- ~~Word Alignment~~
- Expectation Maximization (EM) Algorithm
- IBM Models 1--5
  - IBM Model 1: lexical translation
  - IBM Model 2: alignment model
  - IBM Model 3: fertility
  - IBM Model 4: relative alignment model
  - IBM Model 5: deficiency
- HMM Models: dependent alignment model
- Problems of Word Alignment
- Quality of Word Alignment