Machine Translation
Course 5

Diana Trandabăț

Academic year: 2022-2023

Phrase-based translation

- Phrase-based approach introduced around 1998 by Franz Josef Och & others (Ney, Wong, Marcu)
  - many words to many words (improvement on IBM one-to-many)

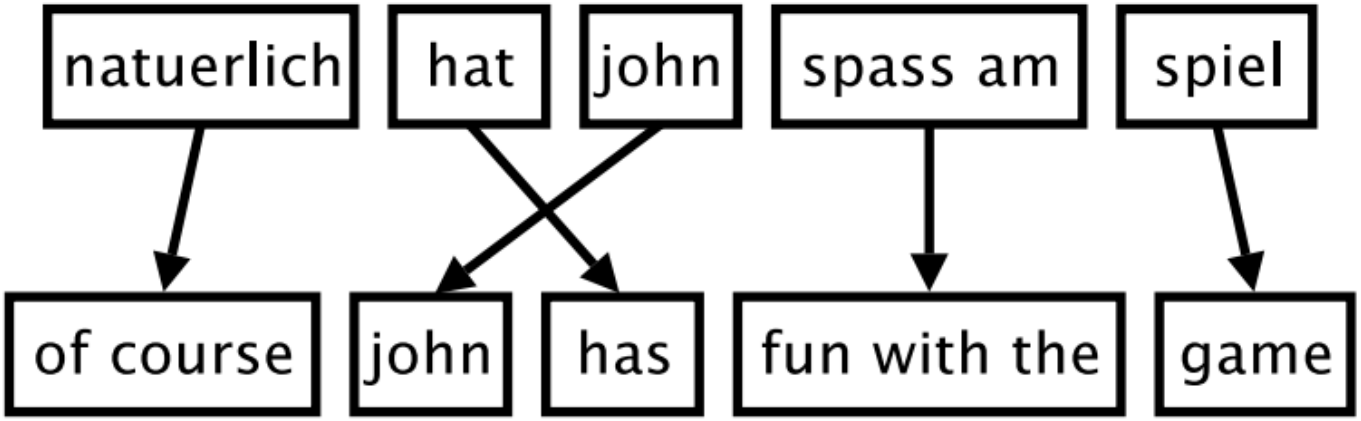
Example: **« cul de sac »**

- word-based translation
  « ass of bag » (N. Am)
  « arse of bag » (British)
- phrase-based translation
  « dead end » (N. Am.)
  « blind alley » (British)

Advantages of phrase-based translation

- Many-to-many translations can handle non-compositional phrases
- Use of local context in translation
- The more data, the longer the phrases which can be learned

Phrase-based translation



- Foreign input is segmented in phrases
- Each phrase is translated into target language
- Phrases are reordered in target language

Phrase Translation Table

- Main knowledge source: table with phrase translations and their probabilities
- Example: phrase translations for *natürlich*

Translation	Probability $\phi(\bar{e}\|\bar{f})$
of course	0.5
naturally	0.3
of course ,	0.15
, of course ,	0.05

Real example

- Phrase translations for *den Vorschlag* learned from the Europarl corpus:

English	$\phi(\bar{e}\|\bar{f})$	English	$\phi(\bar{e}\|\bar{f})$
the proposal	0.6227	the suggestions	0.0114
's proposal	0.1068	the proposed	0.0114
a proposal	0.0341	the motion	0.0091
the idea	0.0250	the idea of	0.0091
this proposal	0.0227	the proposal ,	0.0068
proposal	0.0205	its proposal	0.0068
of the proposal	0.0159	it	0.0068
the proposals	0.0159	...	...

- lexical variation (*proposal* vs. *suggestions*)
- morphological variation (*proposal* vs. *proposals*)
- included function words (*the*, *a*, ...)
- noise (*it*)

Linguistic Phrases?

- Model is not limited to linguistic phrases (noun phrases, verb phrases, prepositional phrases, ...)
- Example of non-linguistic phrase pair

  spass am => *fun with the*

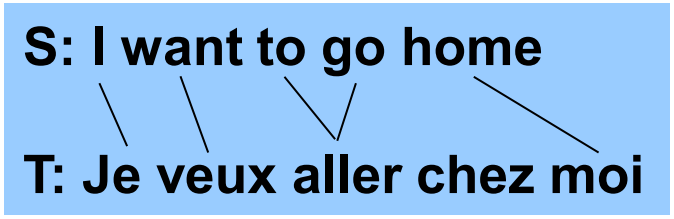
- Prior noun often helps with translation of preposition
- Experiments show that limitation to linguistic phrases hurts quality

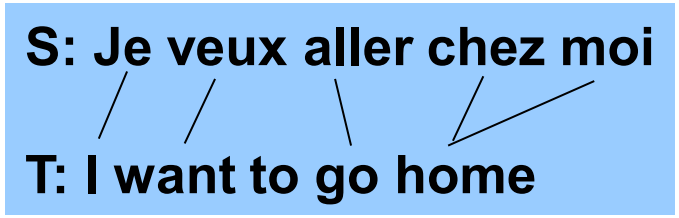
Symmetrized Word Alignment using IBM Models

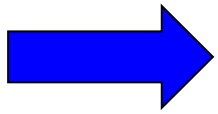
Alignments produced by IBM models are asymmetrical: **source words have at most one connection**, but target words may have many connections.

To improve quality, use symmetrization heuristic :

1. Perform two separate alignments, one in each different translation direction.
2. Take intersection of links as starting point.
3. Add neighbouring links from union until all words are covered.







Phrase Pair Extraction Algorithm

1. Run a sentence aligner on a parallel bilingual corpus
2. Run word aligner (*e.g.*, one based on IBM models) on each aligned sentence pair.
3. From each aligned sentence pair, extract all phrase pairs with no external links (only *consistent* phrase pairs).

Phrase-based probabilistic translation model

- Major components of phrase-based model
  - Phrase translation model $\phi(\mathrm{f}|\mathrm{e})$
  - Reordering model *d*
  - Language model $p_{LM}(e)$
- Bayes rule

$$\begin{aligned} argmax_e p(e|f) &= argmax_e p(f|e) * p(e) \\ &= argmax_e \emptyset(f|e) * p_{LM}(e) \end{aligned}$$

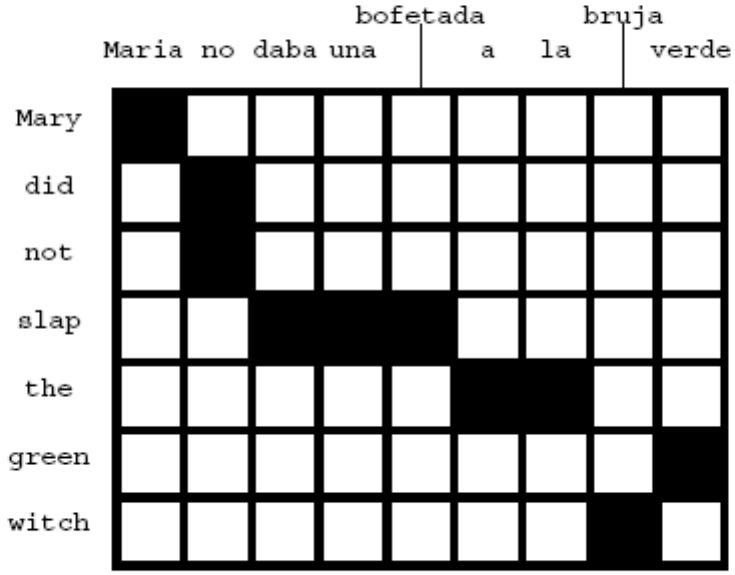
- Sentence in source language is decomposed into *I* phrases $\bar{f}_1^{\,I} = \bar{f}_1, \ldots, \bar{f}_I$
- Decomposition of $\phi(\mathrm{f}|\mathrm{e})$

Learning a Phrase Translation Table

- Task: learn the model from a parallel corpus
- Three stages:
  - word alignment: using IBM models or other method
  - extraction of phrase pairs
  - scoring phrase pairs

How to learn the phrase translation table?

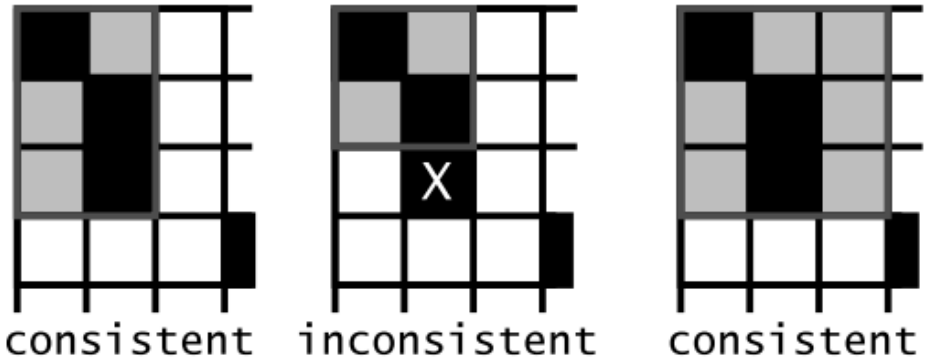
- Start with the *word alignment*:



- Collect all phrase pairs that are **consistent** with the word alignment

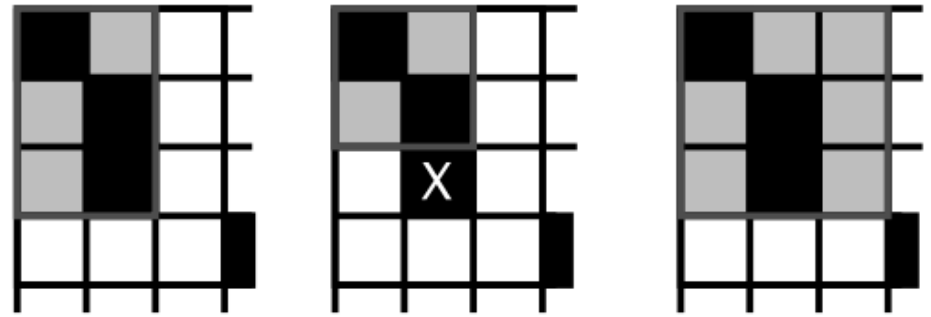
Slide from Koehn 2008

Consistent



All words of the phrase pair have to align to each other.

Consistent



ok

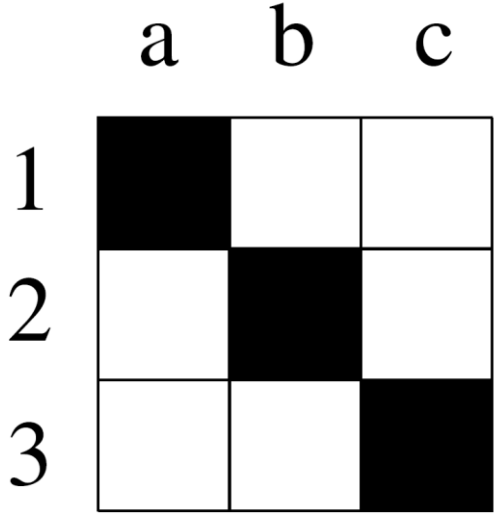
violated
one alignment
point outside

ok
unaligned
word is fine

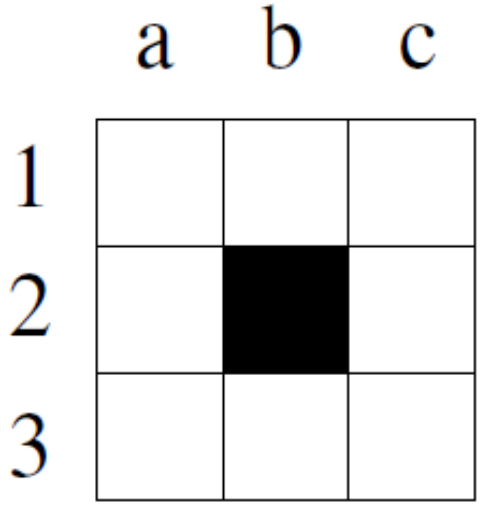
Phrase pair $(\bar{e},\ \bar{f})$ is consistent with an alignment A, if all words $f_1, \ldots, f_n$ in $\bar{f}$ that have alignment points in A have these with words $e_1, \ldots, e_n$ in $\bar{e}$ and vice versa:

$$(\bar{e}, \bar{f})\ consistent\ with\ A \Leftrightarrow$$
$$\forall e_i \in \bar{e} : (e_i, f_j) \in A \rightarrow f_j \in \bar{f}$$
$$AND\ \forall f_j \in \bar{f} : (e_i, f_j) \in A \rightarrow e_i \in \bar{e}$$
$$AND \quad \exists e_i \in \bar{e}, \exists f_j \in \bar{f} : (e_i, f_j) \in A$$

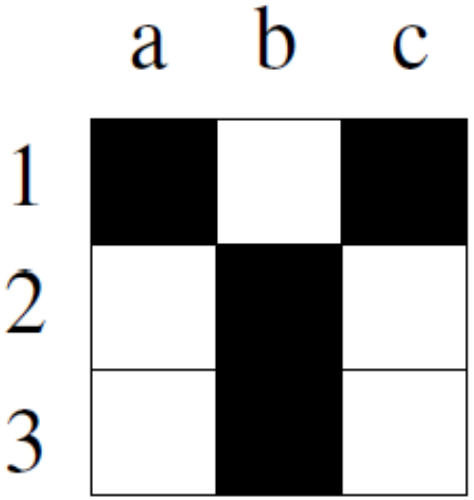
Testing your intuitions



a1, b2, c3, ab12, bc23, abc123

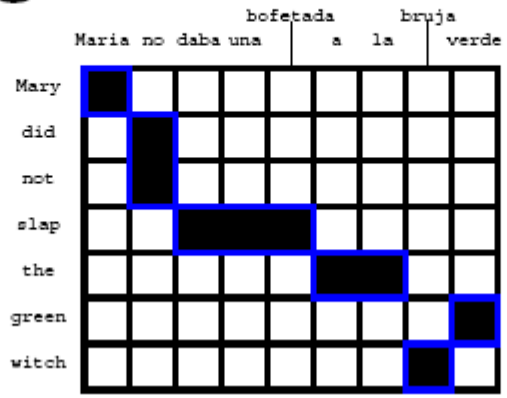


b2, ab12, ab23, ab 123, bc12, bc23, bc123, abc12, abc 23, abc123



b23, abc123

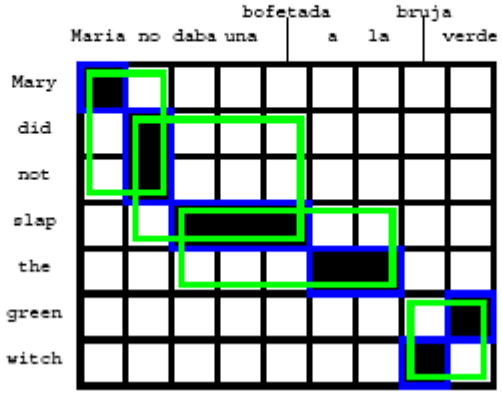
Word alignment induced phrases



(Maria, Mary), (no, did not), (slap, daba una bofetada), (a la, the), (bruja, witch), (verde, green)

Slide from Koehn 2008

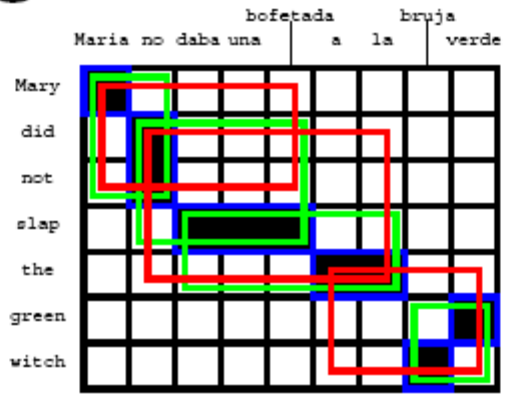
Word alignment induced phrases



(Maria, Mary), (no, did not), (slap, daba una bofetada), (a la, the), (bruja, witch), (verde, green),
(Maria no, Mary did not), (no daba una bofetada, did not slap), (daba una bofetada a la, slap the),
(bruja verde, green witch)

Slide from Koehn 2008

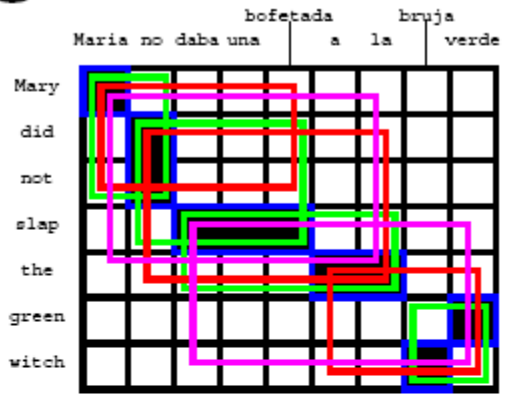
Word alignment induced phrases



(Maria, Mary), (no, did not), (slap, daba una bofetada), (a la, the), (bruja, witch), (verde, green),
(Maria no, Mary did not), (no daba una bofetada, did not slap), (daba una bofetada a la, slap the),
(bruja verde, green witch), (Maria no daba una bofetada, Mary did not slap),
(no daba una bofetada a la, did not slap the), (a la bruja verde, the green witch)

Slide from Koehn 2008

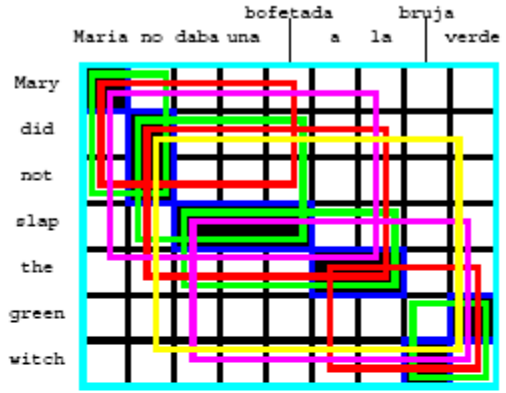
Word alignment induced phrases



(Maria, Mary), (no, did not), (slap, daba una bofetada), (a la, the), (bruja, witch), (verde, green),
(Maria no, Mary did not), (no daba una bofetada, did not slap), (daba una bofetada a la, slap the),
(bruja verde, green witch), (Maria no daba una bofetada, Mary did not slap),
(no daba una bofetada a la, did not slap the), (a la bruja verde, the green witch),
(Maria no daba una bofetada a la, Mary did not slap the),
(daba una bofetada a la bruja verde, slap the green witch)

Slide from Koehn 2008

Word alignment induced phrases (5)



(Maria, Mary), (no, did not), (slap, daba una bofetada), (a la, the), (bruja, witch), (verde, green),
(Maria no, Mary did not), (no daba una bofetada, did not slap), (daba una bofetada a la, slap the),
(bruja verde, green witch), (Maria no daba una bofetada, Mary did not slap),
(no daba una bofetada a la, did not slap the), (a la bruja verde, the green witch),
(Maria no daba una bofetada a la, Mary did not slap the), (daba una bofetada a la bruja verde,
slap the green witch), (no daba una bofetada a la bruja verde, did not slap the green witch),
(Maria no daba una bofetada a la bruja verde, Mary did not slap the green witch)

Slide from Koehn 2008

Probability distribution of phrase pairs

- We need a probability distribution $\emptyset(\bar{f}|\bar{e})$ over the collected phrase pairs
- Possible choices:
  - Relative frequency of collected phrases:

$$\emptyset(\bar{f}|\bar{e}) = \frac{count(\bar{f}|\bar{e})}{\sum_{\bar{f}} count(\bar{f}|\bar{e})}$$

  - Use lexical translation probabilities

Reordering

- Monotone translation
  - Do not allow any reordering
  - Worse translations
- Limiting reordering (movement over max. number of words)
- Distance-based reordering cost
  - Moving a foreign phrase over *n* words: cost z^n
- Lexicalized reordering model

Reordering

- Exercise 1: This task is called bag generation. Put these words in order:
  - “have programming a seen never I language better”.
  - “actual the hashing is since not collision-free usually the is less perfectly the of somewhat capacity table”
- What kind of knowledge are you applying here?
- Do you think a machine could do this job?
- Can you think of a way to automatically test how well a machine is doing, without a lot of human checking?

- Exercise 2. Put these words in order: “loves John Mary”

Great!

See you next time!