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

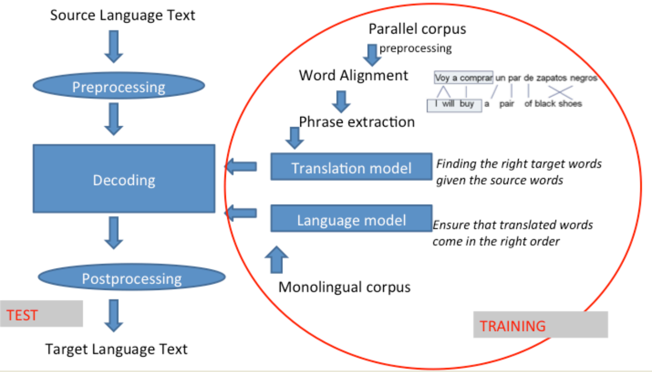
Phrase Extraction and Scoring

Lluís Formiga

July 22, 2014

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

A picture is worth a million equations



Talk Outline

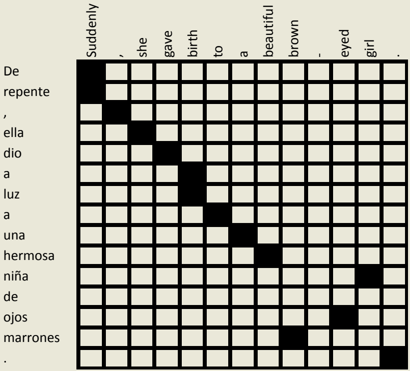
- Phrase-models
  - Alignment
  - Extraction
  - Scoring
  - Log-linear Weighted Model
  - Expanding the Model
- Training Summary

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

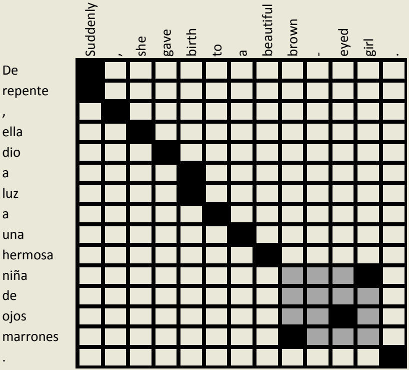
Learning a Phrase Translation Table

- 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

Word Alignment



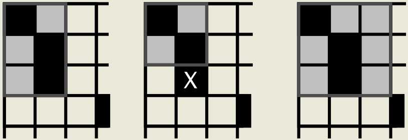
Extracting Phrase Pairs



- extract phrase pair consistent with word alignment:

→ brown-eyed girl / niña de ojos marrones

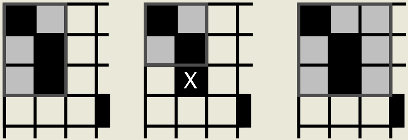
Consistent



consistent	inconsistent	consistent
ok	**violated** one alignment point outside	**ok** unaligned word is fine

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

Consistent



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

$$(\bar{e}, \bar{f}) \text{ consistent with } A \Leftrightarrow$$

$$\forall e_i \in \bar{e} : (e_i, f_j) \in A \rightarrow f_j \in \bar{f}$$

$$\text{AND } \forall f_j \in \bar{f} : (e_i, f_j) \in A \rightarrow e_i \in \bar{e}$$

$$\text{AND } \exists e_i \in \bar{e}, f_j \in \bar{f} : (e_i, f_j) \in A$$

Phrase Extraction Algorithm

- We have the definition of consistency.
- We need an algorithm that extracts all consistent phrase pairs.
- Idea: Loop over all possible e phrases and find the minimal f phrase that matches each of them.
- Matching is done by identifying all alignment points for the e phrase and finding the shortest f phrase that includes all the foreign counterparts for the e words.

Phrase Pair Extraction Considerations

- If the e phrase contains only unaligned words,we do not want to match it against the foreign sentence.
- If the matched minimal foreign phrase has additional alignment points outside the e phrase, we cannot extract this phrase pair. In fact, no phrase pair can be extracted for this e phrase.

Phrase Pair Extraction Considerations

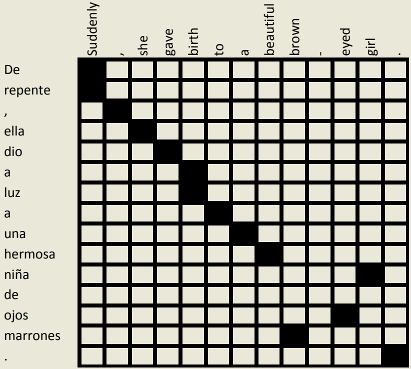
- Other f phrases than the minimally matched f phrase may be consistent with the e phrase. If the f phrase borders unaligned words, then it is extended to these words, and the extended phrase is also added as a translation of the e phrase.

- Role of alignment points: Constraints for which phrase pairs can be extracted. $\Rightarrow$ The fewer alignment points there are, the more phrase pairs can be extracted.

Phrase Pair Extraction Algorithm

```
Input: word alignment A for sentence pair (e,f)
Output: set of phrase pairs BP
for e_start = 1 ... length(e) do
  for e_end = e_start ... length(e) do
    // find the minimally matching foreign phrase
    (f_start, f_end) = ( length(f), 0 )
    for all (e,f) ∈ A do
      if e_start ≤ e ≤ e_end then
        f_start = min( f, f_start )
        f_end = max( f, f_end )
      end if
    end for
    add extract(f_start, f_end, e_start, e_end) to set BP
  end for
end for
function extract(f_start, f_end, e_start, e_end)
return {} if f_end == 0 // check if at least one alignment point
// check if alignment points violate consistency
for all (e,f) ∈ A do
  return {} if e < e_start or e > e_end
end for
// add pharse pairs (incl. additional unaligned f)
E = {}
f_s = f_start
repeat
  f_e = f_end
  repeat
    add phrase pair (e_start .. e_end, f_s .. f_e) to set E
    f_e++
  until f_e aligned
  f_s--
until f_s aligned
return E
```

Phrase Pair Extraction



Smallest phrase pairs:

suddenly — de repente

, — , ; **she — ella** ; **gave — dio**

birth — a luz ; **to — a** ; **a — una**

beautiful — hermosa

girl — niña / niña de

- eyed — de ojos / ojos

eyed — de ojos / ojos

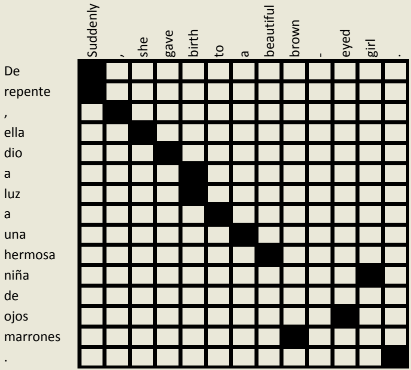
brown — marrones

brown - — marrones

. — .

→ unaligned words (here: Spanish **de** and English **-**) lead to multiple translations

Larger Phrase Pairs



suddenly , — de repente , ; suddenly , she — de repente , ella ; suddenly , she gave — de repente , ella dio ; suddenly , she gave birth — de repente , ella dio a luz ; , she — , ella ; , she gave — , ella dio ; , she gave birth — , ella dio a luz ; , she gave birth to — , ella dio a luz a ; , she gave birth to a — , ella dio a luz a una ; she gave — ella dio ; she gave birth — ella dio a luz ; she gave birth to — ella dio a luz a ; she gave birth to a — ella dio a luz a una ; she gave birth to a beautiful — ella dio a luz a una hermosa ; gave birth — dio a luz ; gave birth to — dio a luz a ; gave birth to a — dio a luz a una ; gave birth to a beautiful — dio a luz a una hermosa ; birth to — a luz a ; birth to a — a luz a una ; birth to a beautiful — a luz a una hermosa ; to a — a una ; to a beautiful — a una hermosa ; to a beautiful brown - eyed girl — a una hermosa niña de ojos marrones ; a beautiful — una hermosa ; a beautiful brown - eyed girl — una hermosa niña de ojos marrones ; a beautiful brown - eyed girl . — una hermosa niña de ojos marrones . ; beautiful brown - eyed girl — hermosa niña de ojos marrones ; beautiful brown - eyed girl . — hermosa niña de ojos marrones . ; eyed girl — niña de ojos ; - eyed girl — niña de ojos ; brown - eyed girl — niña de ojos marrones ; brown - eyed girl . — niña de ojos marrones .

Scoring Phrase Translations

- Phrase pair extraction: collect all phrase pairs from the data
- Phrase pair scoring: assign probabilities to phrase translations
- Score by relative frequency:

$$\phi(\bar{f}|\bar{e}) = \frac{\text{count}(\bar{e}, \bar{f})}{\sum_{\bar{f}_i} \text{count}(\bar{e}, \bar{f}_i)}$$

Size of the Phrase Table

- Phrase translation table typically bigger than corpus

  ... even with limits on phrase lengths (e.g., max 7 words)

→ Too big to store in memory?

- Solution for training
  - extract to disk, sort, construct for one source phrase at a time
- Solutions for decoding
  - on-disk data structures with index for quick look-ups
  - suffix arrays to create phrase pairs on demand

Weighted Model

- Described standard model consists of three sub-models
  - phrase translation model $\phi(\bar{f}|\bar{e})$
  - reordering model d
  - language model $p_{LM}(e)$

$$e_{best} = \text{argmax}_e \prod_{i=1}^{I} \phi(\bar{f}_i|\bar{e}_i)\ d(start_i - end_{i-1} - 1) \prod_{i=1}^{|e|} p_{LM}(e_i|e_1 \ldots e_{i-1})$$

- Some sub-models may be more important than others
- Add weights λ_ϕ, λ_d, λ_{LM}

$$e_{best} = \text{argmax}_e \prod_{i=1}^{I} \phi(\bar{f}_i|\bar{e}_i)^{\lambda_\phi}\ d(start_i - end_{i-1} - 1)^{\lambda_d} \prod_{i=1}^{|e|} p_{LM}(e_i|e_1 \ldots e_{i-1})^{\lambda_{LM}}$$

Log-Linear Model

- Such a weighted model is a log-linear model:

$$p(x) = \exp \sum_{i=1}^{n} \lambda_i h_i(x)$$

- Our feature functions
  - number of feature function $n = 3$
  - random variable $x = (e, f, start, end)$
  - feature function $h_1 = \log \phi$
  - feature function $h_2 = \log d$
  - feature function $h_3 = \log p_{LM}$

Weighted Model as Log-Linear Model

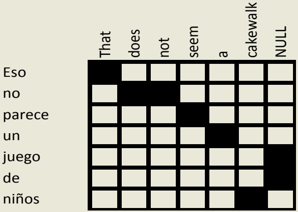
$$p(e, a|f) = \exp(\lambda_\phi \sum_{i=1}^{I} \log \phi(\bar{f}_i|\bar{e}_i) +$$

$$\lambda_d \sum_{i=1}^{I} \log d(a_i - b_{i-1} - 1) +$$

$$\lambda_{LM} \sum_{i=1}^{|e|} \log p_{LM}(e_i|e_1 \ldots e_{i-1}))$$

More Feature Functions

- Bidirectional alignment probabilities: $\phi(\bar{e}|\bar{f})$ and $\phi(\bar{f}|\bar{e})$
- Rare phrase pairs have unreliable phrase translation probability estimates
  $\rightarrow$ lexical weighting with word translation probabilities



$$\text{lex}(\bar{e}|\bar{f}, a) = \prod_{i=1}^{\text{length}(\bar{e})} \frac{1}{|\{j|(i,j) \in a\}|} \sum_{\forall (i,j) \in a} w(e_i|f_j)$$

More Feature Functions

- Language model has a bias towards short translations

  $\rightarrow$ word count: $\mathrm{wc}(\mathrm{e}) = \log |\mathrm{e}|^{\omega}$

- We may prefer finer or coarser segmentation

  $\rightarrow$ phrase count $\mathrm{pc}(\mathrm{e}) = \log |\mathrm{I}|^{\rho}$

- Multiple language models
- Multiple translation models
- Other knowledge sources

EM Training of the Phrase Model

- We presented a heuristic set-up to build phrase translation table
  (word alignment, phrase extraction, phrase scoring)
- Alternative: align phrase pairs directly with EM algorithm
  - initialization: uniform model, all $\phi(\bar{e}, \bar{f})$ are the same
  - expectation step:
    - estimate likelihood of all possible phrase alignments for all sentence pairs
  - maximization step:
    - collect counts for phrase pairs $(\bar{e}, \bar{f})$, weighted by alignment probability
    - update phrase translation probabilties $p(\bar{e}, \bar{f})$
- However: method easily overfits
  (learns very large phrase pairs, spanning entire sentences)

Summary

- Lexical Translation / Alignment and Reordering
- Phrase Model
- Training the model
  - word alignment
  - phrase pair extraction
  - phrase pair scoring
- Log linear model
  - sub-models as feature functions
  - lexical weighting
  - word and phrase count features
- EM training of the phrase model