INF5820/INF9820

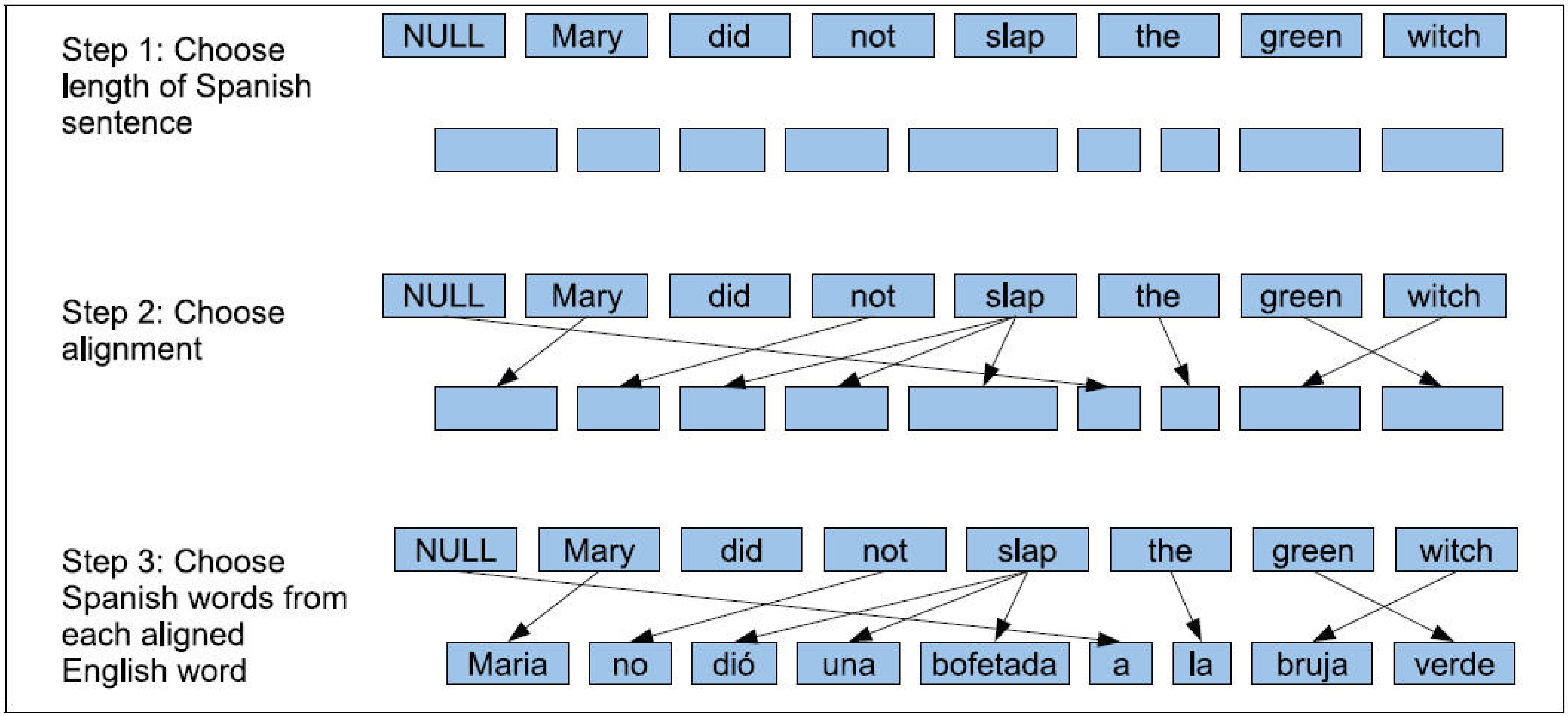
LANGUAGE TECHNOLOGICAL APPLICATIONS

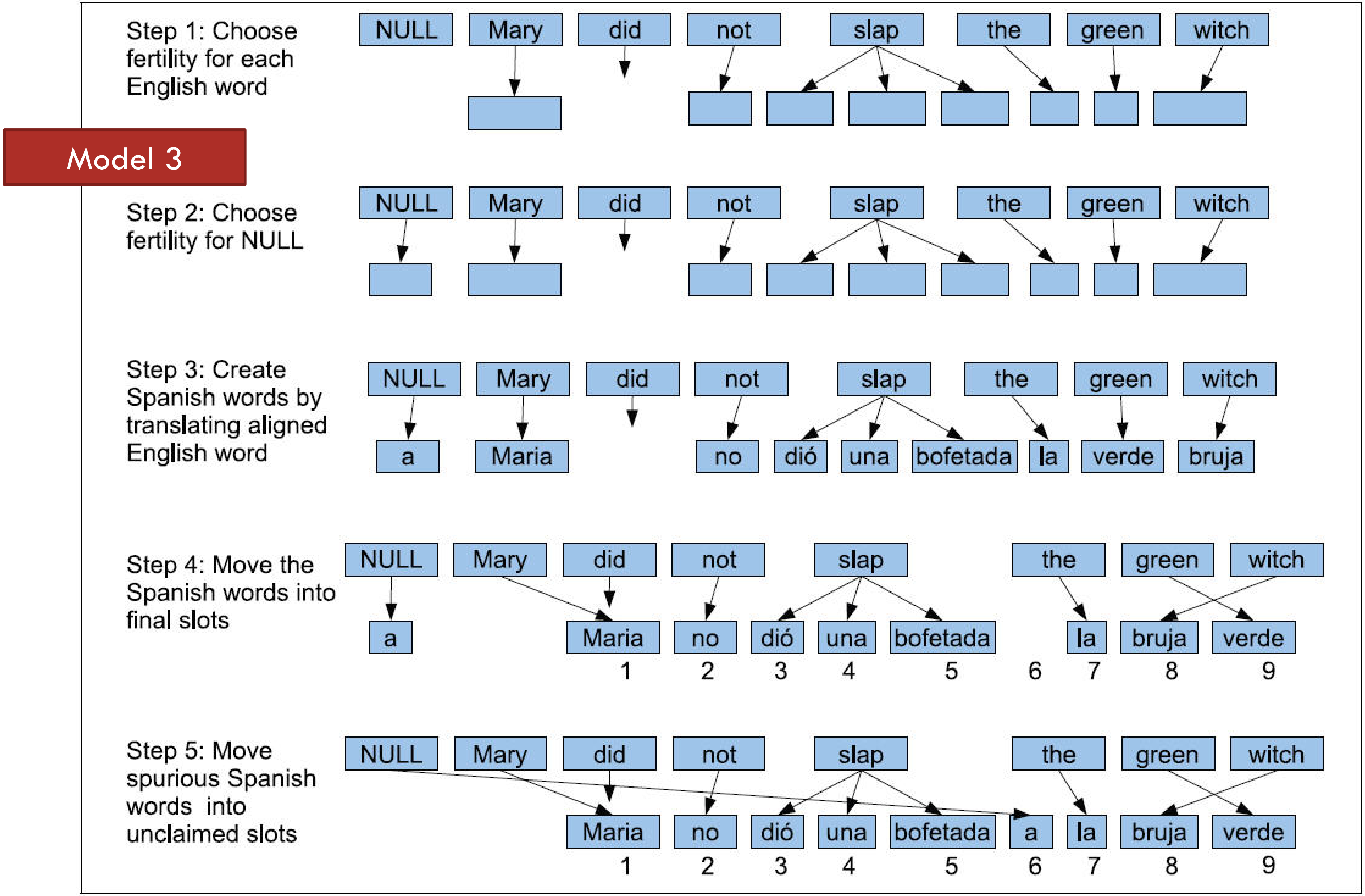
Jan Tore Lønning, Lecture 6, 28 Sep. 2016
jtl@ifi.uio.no

Today

- Repetisjon:
  - Higher IBM-models: 3, 4, 5
- Phrase-Based Models

Model 1 & 2 and HMM alignment




Model 3
Step 1: Choose fertility for each English word
NULL
Mary
did
not
slap
the
green
witch
Step 2: Choose fertility for NULL
Step 3: Create Spanish words by translating aligned English word
a
Maria
no
dió
una
bofetada
la
verde
bruja
Step 4: Move the Spanish words into final slots
1
2
3
4
5
6
7
8
9
Step 5: Move spurious Spanish words into unclaimed slots

IBM Model 3: Fertility

- Fertility: number of F words produced by an E word
- Modelled by a distribution $n(x|e)$

Example:
F = Norw.
n(2 | yesterday) ≈ 1
n(1 | to) ≈ 0.8
n(2 | to) ≈ 0.2
n(1 | car) ≈ 1
n(0 | the) ≈ 0.6
n(1 | the) ≈ 0.4

Example:
Norw.→Eng.
n(2 | bilen) ≈ 0.7
n(1 | bilen) ≈ 0.3
n(1 | å) ≈ 0.8
n(0 | å) ≈ 0.2

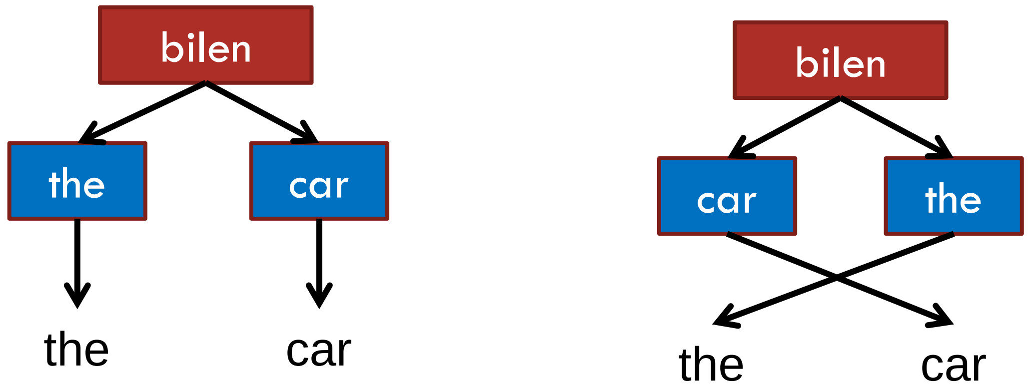
IBM Model 3: Null insertion

- Modelled by:
- There is a probability p0:
  - After each inserted word there is the probability p0 of not inserting a null-word
  - And a probability p1 = (1-p0) of inserting a null-word
- A rather complex expression for what this contributes into P(a, **f**|**e**) which considers
  - Permutations
  - Length of **f**

IBM Model 3: Distortion

$$d(j \mid a_j, m, k)$$

- A probability distribution which gives the probability of word a_i ending up in position j.
- Similar to alignment in model 2 but:
  - Opposite direction
  - Different choices of words + distortion may correpsond to the same alignment



IBM model 3

$$P(\mathbf{f},\mathbf{a}\,|\,\mathbf{e}) = \prod_{j=1}^{m} t(f_j \,|\, e_{a_j}) \prod_{j=1}^{m} d(j \,|\, a_j, k, m) \times \text{ more}$$

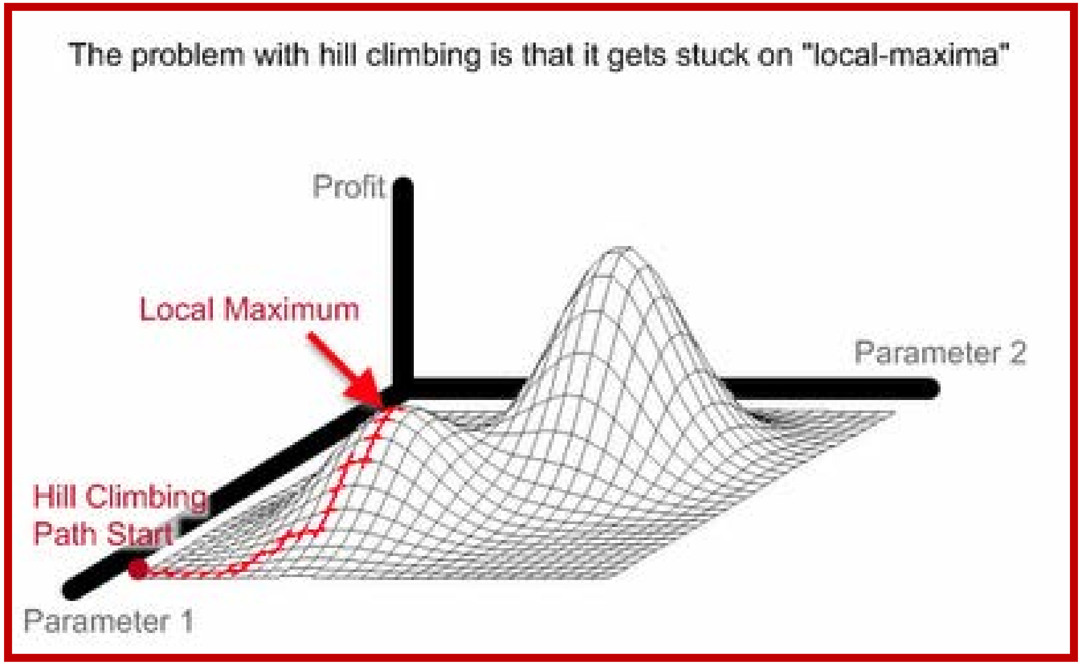
- Where *more* is an expression which counts
  - $n(x\,|\,e_i)$ the right number of times
  - And uses p0 to give the right probability to null-insertion.

Training Model 3

- In principle like Model 1, but
  - The trick to get rid of the alignments does not work
  - Too costly to calculate all alignments
- Strategy
  - Sample and use the most probable alignments
  - Start with alignments from Model 1 and Model 2
  - Use hill-climbing algorithm

Hill-climbing algorithm

- Assign some initial parameter values
- Consider (a slightly different) value for <u>one</u> of the parameters and see whether the result is better:
  - If YES, change the parameter accordingly
- Repeat
  - (until we do not see big improvements).



Training model 3

- Model 1: The optimum we find is global
- Model 3 (and model 2):
  - A local optimum does not have to be global
- First run some iterations of Model1 and maybe some iterations of Model 2
- Use the results, in particular the alignment, as input to Model 3
- Hill-climb the space of alignments from here, doing minimal changes.

IBM Model 4

- Better reordering model
- Consider group of words (phrases)
- Distinguish between
  - the placement of the whole group
  - The placement within the group

The IBM-models

- IBM models 1-4 are not true probability models.
- Model 5 fixes this
  - Based of model 4
- We will not consider models 4 and 5
- Phrase-Based translation makes use of Model 3

Today

- Higher IBM-models: 3, 4, 5
- Phrase-Based Models

Phrase alignment

- K. Slides to chapter 4:
  - 49-51
  - 53-54

Phrase-Based Models

- K. Slides to chapter 5