CSCI 5832
Natural Language Processing

Lecture 26
Jim Martin

Today 4/26

- **Review**
  - **Machine translation framework**
  - **Word-based models**
- **Worked out EM alignment example**
- **Phrase-based translation**
- **Syntax-based…**

Sample from Today's Al Jazeera

بفوزها بـ 172 مقعدا من أصل 250

الجبهة التقدمية تحصد معظم مقاعد البرلمان السوري

سيارات أحد المرشحين المستقلين الفائزين بمقعد البرلمان السوري تطوف دمشق بعد إعلان النتيجة (فرنسية)

فازت الجبهة الوطنية التقدمية التي يقودها حزب البعث العربي الاشتراكي الحاكم بأغلبية الأصوات في الانتخابات التشريعية السورية التي جرت مطلع الأسبوع الجاري.

وقال وزير الداخلية السوري بسام عبد المجيد إن الجبهة -ائتلاف مكون من عشر أحزاب يقودها البعث- حصلت على 172 مقعدا من أصل 250 بمجلس الشـ السوري، فيما فاز المرشحون المستقلون بـ78 مقعدا.وأشار عبد المجيد في مؤتمر صحفي إلى أن نسبة المشاركة بلغت 56٪.

وأكد الوزير السوري أن أكثر من 50٪ من مقاعد البرلمان المخصصة للجبهة التقدمية ذهبت للعمال والفلاحين الذين يشكلون العمود الفقري لحزب البعث، فيما ذهبت المقاعد المتبقية لبقية فئات الشعب. وحافظت النساء على نفس المقاعد التي حصلن عليها في الدورة السابقة وهو 30 مقعدا.

وأضاف أن البرلمان الجديد يضم 180 عضوا جديدا و70 عضوا حافظوا على مقاعدهم في المجلس.ونفى الوزير وقوع أي مخالفات أمنية خلال مجريات الانتخاب التي سادها "جو آمن وبدون حوادث".

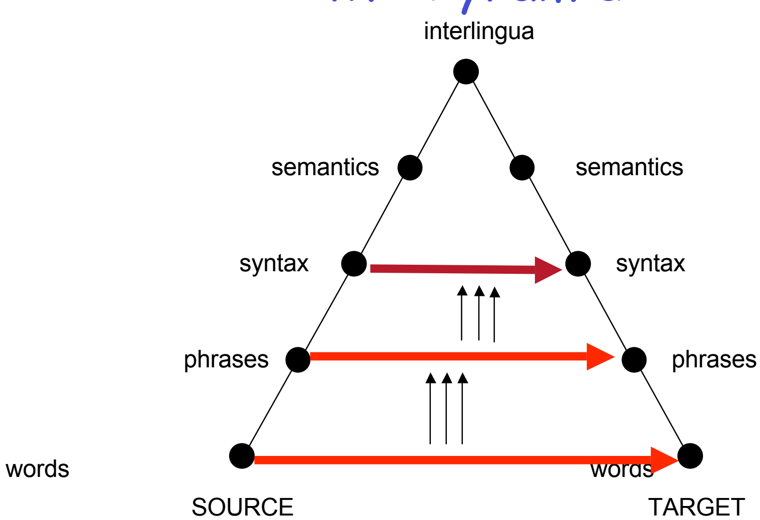
Progressive National Front, led by the Arab Baath Socialist Party a majority of votes in the legislative elections that took place Syrian early this week.

Syrian Interior Minister Bassam Abdul Majid said the Front-a coalition of parties led by the Baath ten-got 172 seats out of 250 to the Syrian people, while independent candidates won 78 seats. He Abdel Meguid said in a press conference that the percentage of participation reached 56%.

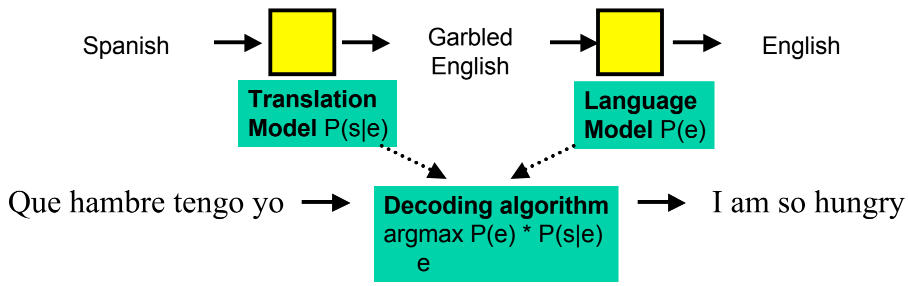
The Syrian minister affirmed that more than 50% of the seats in Parliament for Progressive Front went to the workers and peasants, who form the backbone of the Baath Party, while the remaining seats went to other classes of people. The women maintained the same seats which they had received in the previous session of the 30 seats.

He added that the new parliament includes 180 new members and 70 members retained their seats in the Council. The minister denied any violations of the law during the course of the elections which was overwhelmed by the "atmosphere of security and without incident."

MT Pyramid



Bayes Rule/Noisy Channel



Given a source sentence s, the decoder should consider many possible translations … and return the target string e that maximizes

P(e | s)

By Bayes Rule, we can also write this as:

P(e) x P(s | e) / P(s)

and maximize that instead. P(s) never changes while we compare different e's, so we can equivalently maximize this:

P(e) x P(s | e)

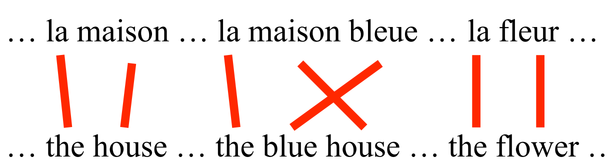
Parts List

- **We need probabilities for**
  - **n (x|y) The probability that word y will yield x outputs in the translation… (fertility)**
  - **p The probability of a null insertion**
  - **t The actual word translation probability table**
  - **d(j|i) the probability that a word at position i will make an appearance at position j in the translation**

Parts List

- Every one of these can be learned from a sentence aligned corpus…
  - Ie. A corpus where sentences are paired but nothing else is specified
- And the EM algorithm

Word Alignment



Inherent hidden structure revealed by EM training!

EM: Worked out example

- **Focus only on the word translation probs**

… la maison … la maison bleue … la fleur …

… the house … the blue house … the flower …

- **How many alignments are there?**

EM: Worked out Knight example

- **Focus only on the word translation probs**

… la maison … la maison bleue … la fleur …

… the house … the blue house … the flower …

2 6 2

- **How many alignments are there?**

EM: Step 1

- Make up some numbers for the parameters of interest. In this case, just the word translation probabilities.

(la|the) (m|the) (b|the) (f|the)	(la|house) (m|house) (b|house)	(la|blue) (m|blue) (b|blue)	(la|flower) (f|flower)

EM: Step 1

- Make up some numbers for the parameters of interest. In this case, just the translation probs.

word translation probs

(la|the) 1/4 (m|the) 1/4 (b|the) 1/4 (f|the) 1/4	(la|house) 1/3 (m|house) 1/3 (b|house) 1/3	(la|blue) 1/3 (m|blue) 1/3 (b|blue) 1/3	(la|flower) 1/2 (f|flower) 1/2

EM: Step 2

- **Make some simple assumptions and produce some normalized alignment probabilities**

la maison
the house

la maison
the house

la maison bleue
the blue house

¼ * 1/3 = 1/12 ¼ * 1/3 = 1/12 ¼ * 1/3 * 1/3 = 1/36

EM: Step 2 (normalize)

- **Make some simple assumptions and produce some normalized alignment probabilities**

la maison
the house

la maison
the house

la maison bleue
the blue house

¼ * 1/3 = 1/12 ¼ * 1/3 = 1/12 ¼ * 1/3 * 1/3 = 1/36

(1/12) / (2/12) = 1/2 (1/12) / (2/12) = 1/2 1/6 For each

EM: Step 3

- **Now that we have the probability of each alignment we can go back and count the evidence in the alignments for each translation pair and prorate them based on the alignments they come from.**

EM: Step 3

- **Now that we have the probability of each alignment we can go back and count the evidence in the alignments for each translation pair and prorate them based on the alignments they come from. Huh?**
- **Let's just look at (la | the).**
  - **What evidence do we have?**

EM: Step 3

- **Evidence for (la|the)**

… la maison … la maison bleue … la fleur …

… the house … the blue house … the flower …

1 2 1

EM: Step 3

- **Evidence for (la|the)**

… la maison … la maison bleue … la fleur …

… the house … the blue house … the flower …

1 2 1

1* ½ 2* 1/6 1* ½

EM: Step 3

- **Evidence for (la|the)**

… la maison … la maison bleue … la fleur …

… the house … the blue house … the flower …

1 2 1

1* ½ 2* 1/6 1* ½

8/6 = Discounted count for (la|the)

EM: Step 3

- **Do that for the other (?|the) and normalize**

(la| the) = 8/6	(la| the) = 8/6 / 18/6	(la| the) = .44
(m| the) = 5/6	(m| the) = 5/6 / 18/6	(m| the) = .27
(b| the) = 2/6	(b| the) = 2/6 /18/6	(b| the) = .11
(f | the) = 3/6	(f | the) = 3/6 / 18/6	(f | the) = .16

EM: Step 4

- Do that for all the words in the table
- Recompute the alignment probabilities using the new word translation probabilities
- Recompute the fractional counts
  - Normalize to get new word translation probs
- Iterate until done
- When you're done you can use the numbers to get the most probable alignment. From which you can read off all the parameters you need.

EM/Alignment

- This is totally weird voodoo science
- We need some parameters
  - which we don't have
- We can get them from a word-aligned corpus
  - which we don't have
- So we make up some parameters to get the alignment and then use that alignment to get the right numbers.

Parts List

- Given a sentence alignment we can induce a word alignment
- Given that word alignment we can get the p, t, d and n parameters we need for the model.
- Ie. We can argmax P(e|f) by max over P(f|e)*P(e)… and we can do that by iterating over some large space of f possibilities.

Break
Quiz Coverage

- Semantics (Compositional and IE)
  - Chapter 16
    - Pages 1-29
  - Chapter 19
    - Sections 19.1-19.4
  - Review Bioinformatics Slides
- Discourse
  - Chapter 21
    - Pages 1-31 (skip 20.6.2)
- MT
  - Chapter 24
    - Pages 1-41 (skip 24.2)

Decoding

- **Remember Viterbi? Just a fancier Viterbi**
  - *Given foreign sentence f, find English sentence e that maximizes P(e) x P(f | e)*
  - *Space is defined by the model (fertility, distortion, word translation model, etc.)*
  - *Large space --> efficient decoding is required.*

Decoding

Que	hambre	tengo	yo
what	hunger	have	I
that	hungry	am	me
so		make	
where			

Decoding

Que	hambre	tengo	yo
what	hunger	have	I
that	hungry	am	me
so		make	
where			

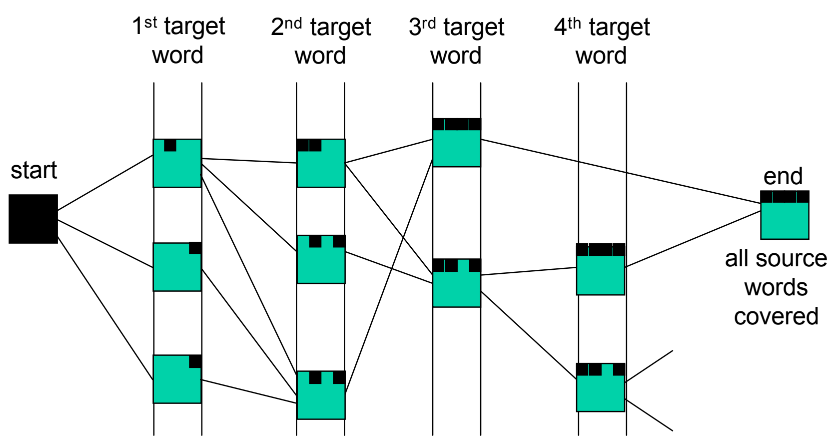
Decoding

Que	hambre	tengo	yo
what	hunger	have	I
that	hungry	am	me
so		make	
where			

Decoding

Que	hambre	tengo	yo
what	hunger	have	I
that	hungry	am	me
so		make	
where			

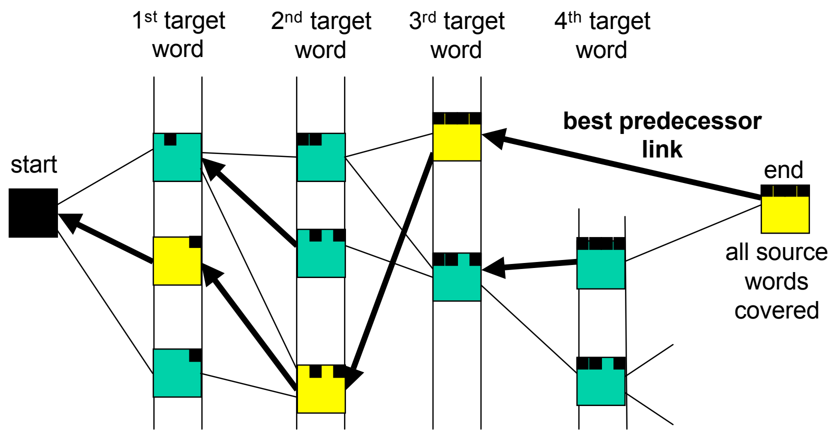
Decoder: Actually Translates New Sentences



Each partial translation hypothesis contains:
- Last English word chosen + source words covered by it
- Entire coverage vector (so far) of source sentence
- Language model and translation model scores (so far)

[Jelinek, 1969;
Brown et al, 1996 US Patent;
(Och, Ueffing, and Ney, 2001]

Dynamic Programming Beam Search



Each partial translation hypothesis contains:
- Last English word chosen + source words covered by it
- Entire coverage vector (so far) of source sentence
- Language model and translation model scores (so far)

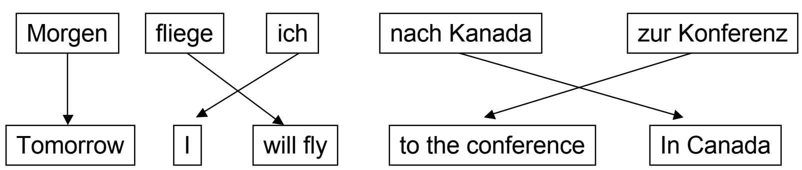
[Jelinek, 1969; Brown et al, 1996 US Patent; (Och, Ueffing, and Ney, 2001]

Flaws of Word-Based MT

- Multiple English words for one French word
  - IBM models can do one-to-many (fertility) but not many-to-one
- Phrasal Translation
  - "real estate", "note that", "interest in"
- Syntactic Transformations
  - Languages with differing word orders (SVO vs. VSO)
  - Translation model penalizes any proposed re-ordering
  - Language model not strong enough to force the verb to move to the right place

Phrase-Based Statistical MT

Phrase-Based Statistical MT



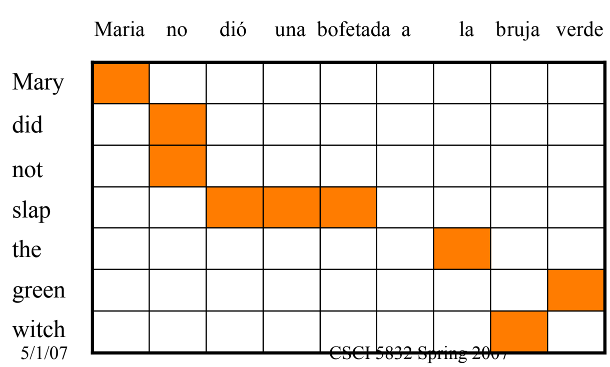
- **Source (foreign) input segmented in to phrases**
  - **"phrase" is any sequence of words**
- **Each phrase is probabilistically translated into English**
  - **P(to the conference | zur Konferenz)**
  - **P(into the meeting | zur Konferenz)** **HUGE TABLE!!**
- **Phrases are probabilistically re-ordered**

Advantages of Phrase-Based

- **Many-to-many mappings can handle non-compositional phrases (e.g., "real estate")**
- **Local context is very useful for disambiguating**
  - **"Interest rate" → …**
  - **"Interest in" → …**
- **The more data, the longer the learned phrases**
  - **Sometimes whole sentences**
    - **Interesting parallel to concatenative synthesis for TTS**

How to Learn the Phrase Translation Table?

- **One method: "alignment templates"** (Och et al, 1999)
- **Start with word alignment, build phrases from that.**

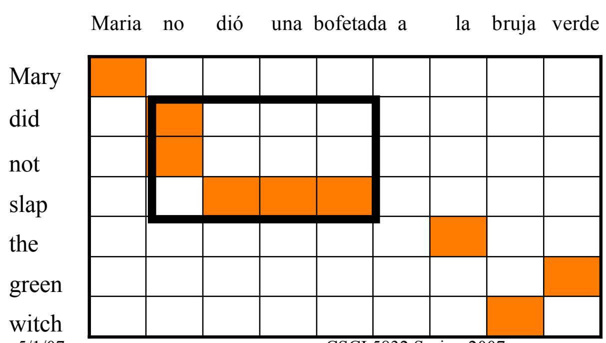


This word-to-word alignment is a by-product of training a translation model like IBM-Model-3.

This is the best (or "Viterbi") alignment.

How to Learn the Phrase Translation Table?

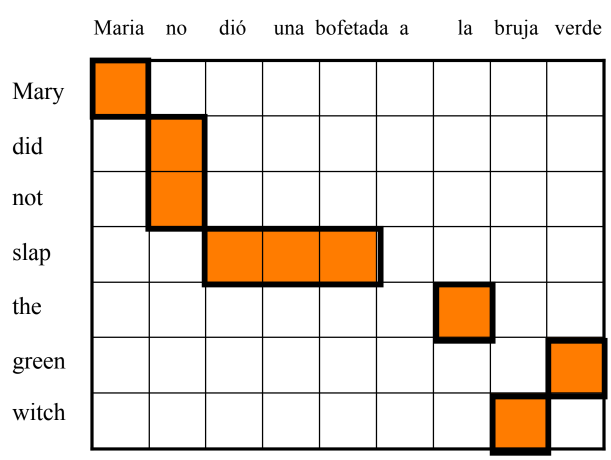
- **One method: "alignment templates"** (Och et al, 1999)
- **Start with word alignment, build phrases from that.**



This word-to-word alignment is a by-product of training a translation model like IBM-Model-3.

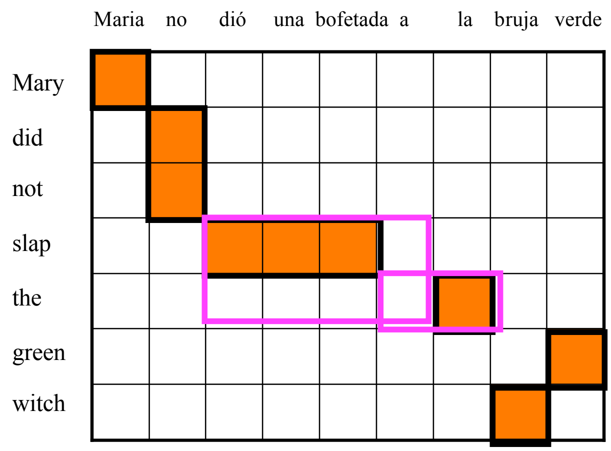
This is the best (or "Viterbi") alignment.

Word Alignment Induced Phrases



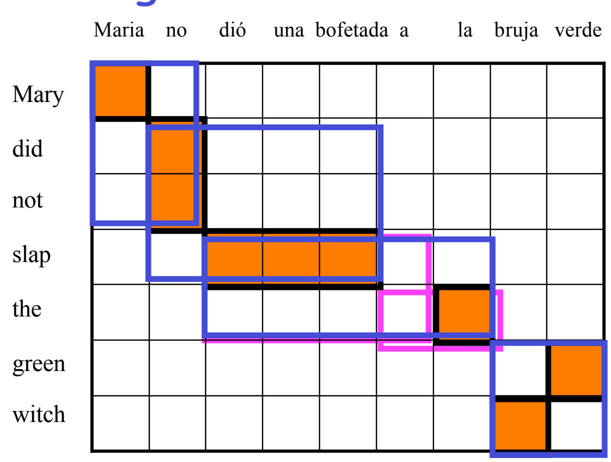
(Maria, Mary) (no, did not) (slap, dió una bofetada) (la, the) (bruja, witch) (verde, green)

Word Alignment Induced Phrases



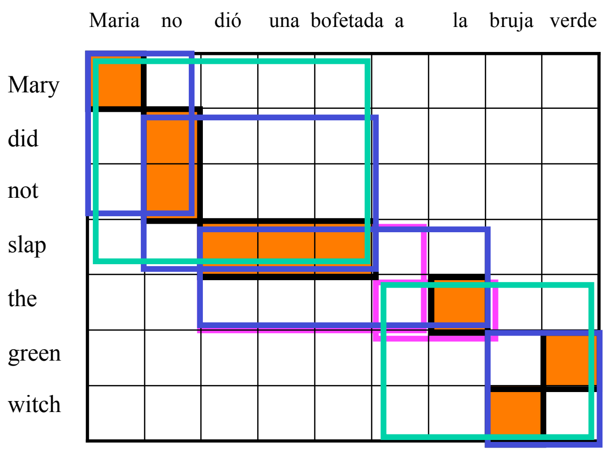
(Maria, Mary) (no, did not) (slap, dió una bofetada) (la, the) (bruja, witch) (verde, green)
(a la, the) (dió una bofetada a, slap the)

Word Alignment Induced Phrases



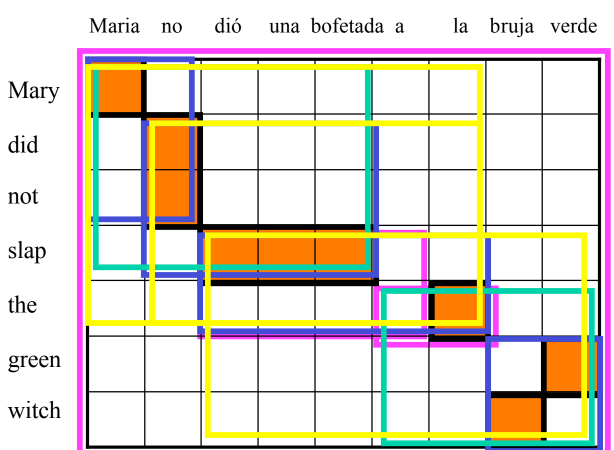
(Maria, Mary) (no, did not) (slap, dió una bofetada) (la, the) (bruja, witch) (verde, green)
(a la, the) (dió una bofetada a, slap the)
(Maria no, Mary did not) (no dió una bofetada, did not slap), (dió una bofetada a la, slap the)
(bruja verde, green witch)

Word Alignment Induced Phrases



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

Word Alignment Induced Phrases



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

Phrase Pair Probabilities

- **A certain phrase pair (f-f-f, e-e-e) may appear many times across the bilingual corpus.**
  - **We hope so!**
- **We can calculate phrase substitution probabilities P(f-f-f | e-e-e)**
- **We can use these in decoding**
- **Much better results than word-based translation!**