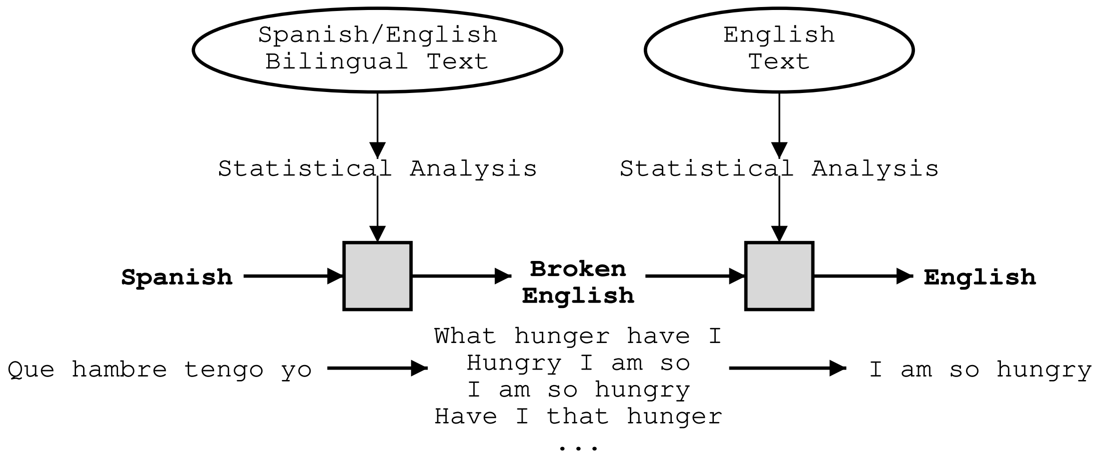
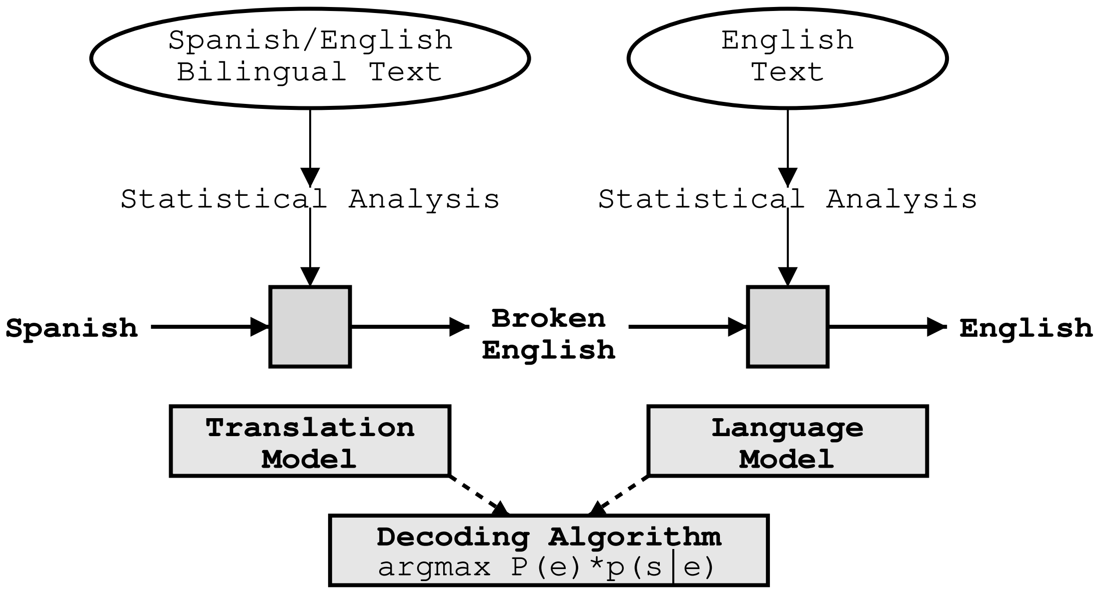
Outline

- **Introduction to Statistical Machine Translation**
- Translation Model: IBM Models and the EM Algorithm
- Language Model
- Decoding Algorithm
- Phrase-Based Machine Translation
- Log-Linear Models
- The Challenge of Syntax
- Divide and Conquer

Statistical MT Systems



Statistical MT Systems (2)



Three Problems in Statistical MT

- Translation Model
  - given a pair of strings $< f, e >$, assigns $P(f|e)$ by formula
  - $< f, e >$ look like translations $\Rightarrow$ high $P(f|e)$
  - $< f, e >$ don't look like translations $\Rightarrow$ low $P(f|e)$
- Language Model
  - given an English string e, assigns $P(e)$ by formula
  - good English string $\Rightarrow$ high $P(e)$
  - bad English string $\Rightarrow$ low $P(e)$
- Decoding Algorithm
  - given a language model, a translation model and a new sentence f, find translation e maximizing $P(e) \times P(f|e)$

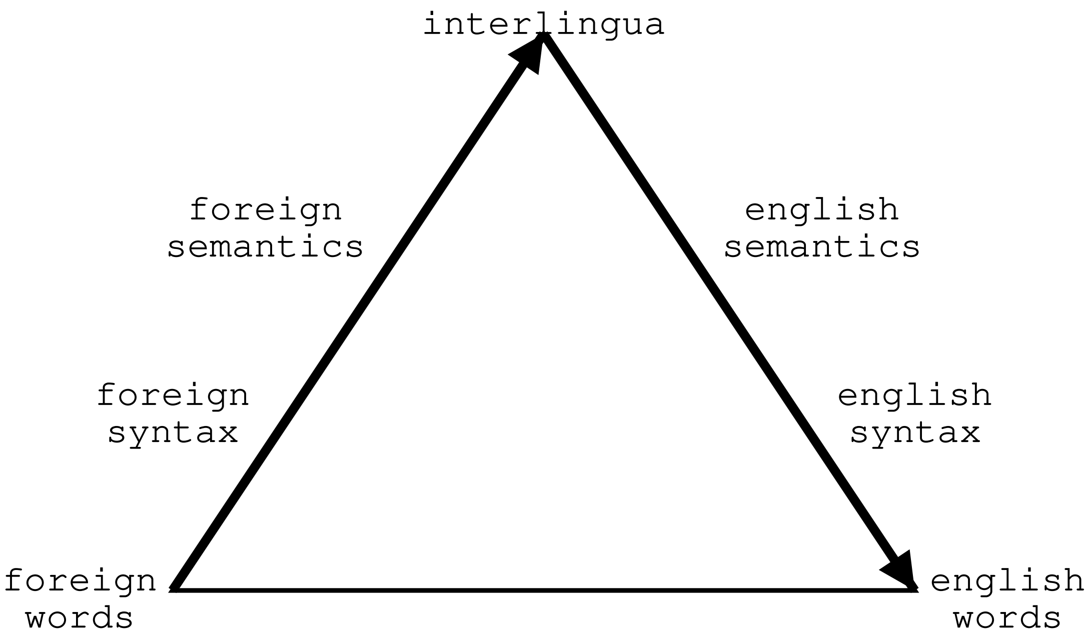
Outline

- Introduction to Statistical Machine Translation
- **Translation Model: IBM Models and the EM Algorithm**
- Language Model
- Decoding Algorithm
- Phrase-Based Machine Translation
- Log-Linear Models
- The Challenge of Syntax
- Divide and Conquer

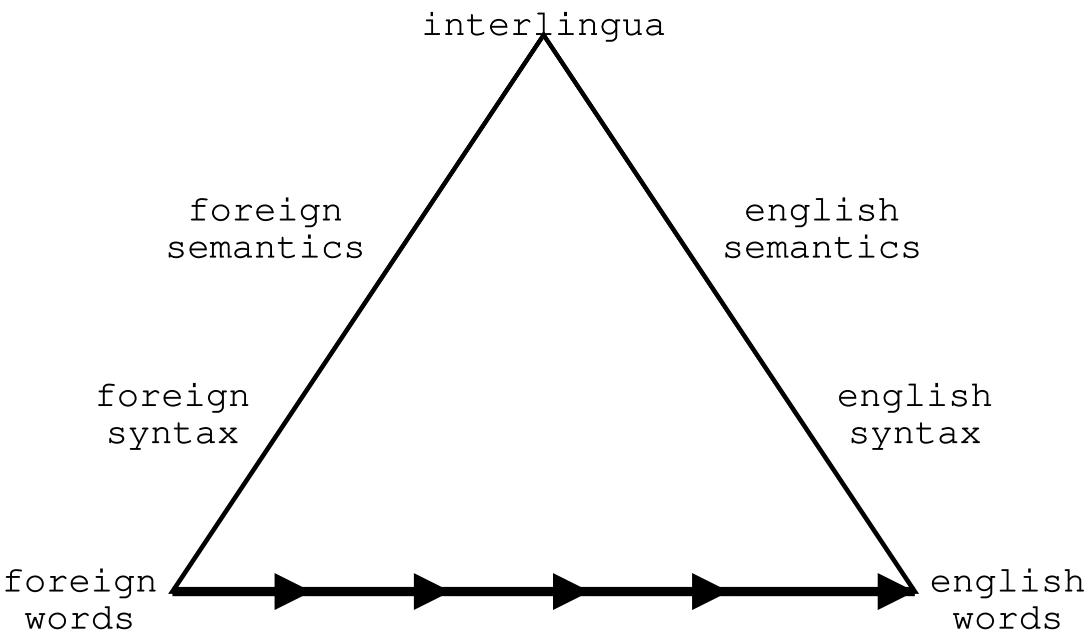
Translation Model

- Goal of the Translation Model:
  match foreign input to English output

The Machine Translation Pyramid

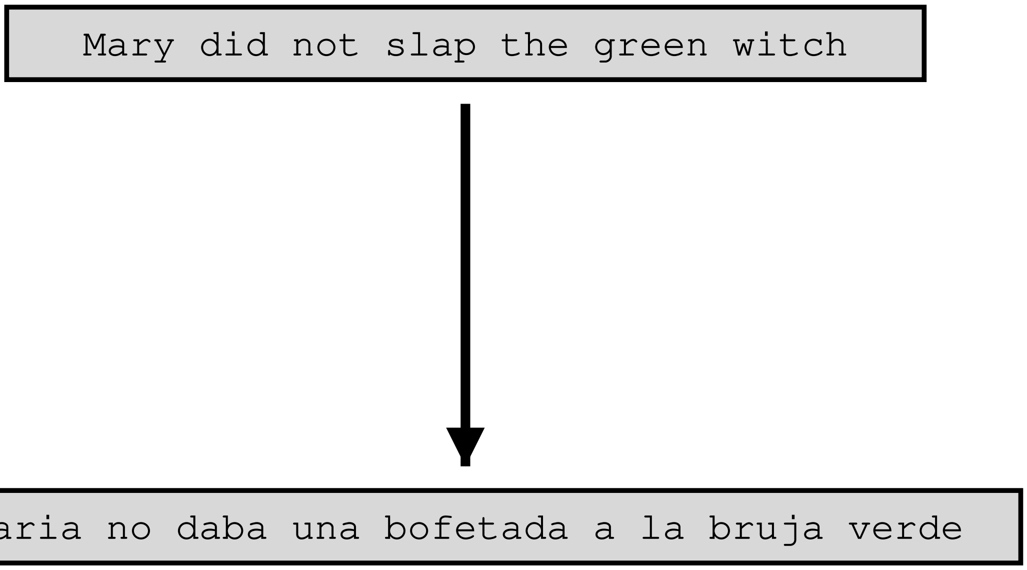


The Machine Translation Pyramid



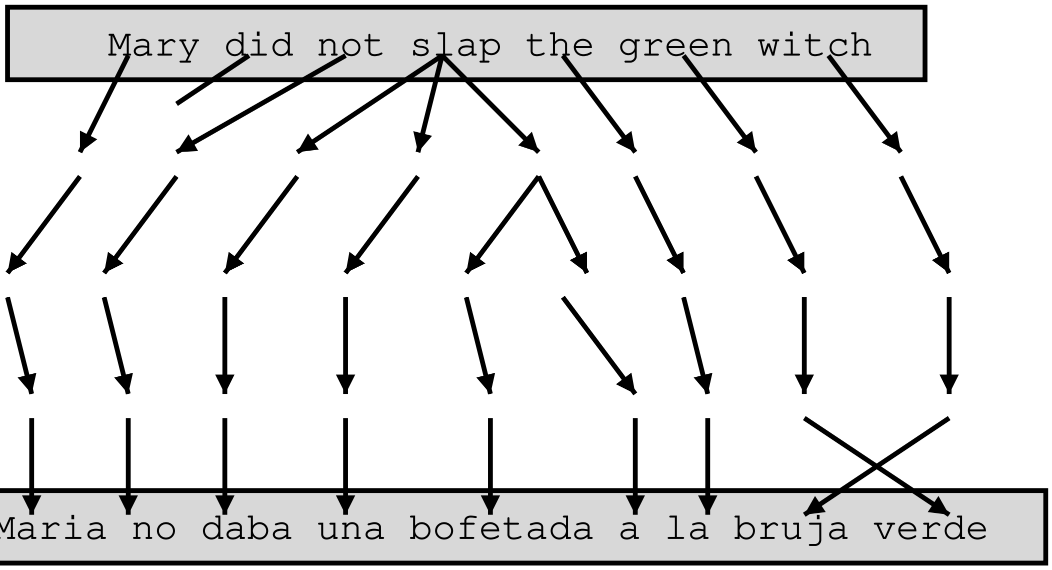
however, the currently best performing statistical machine translation systems are still crawling at the bottom.

Statistical Modeling



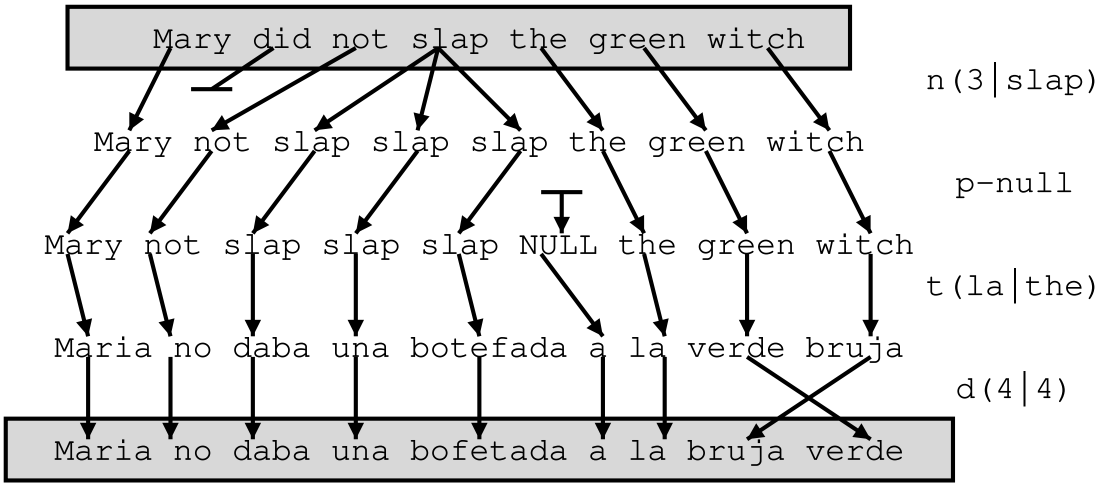
- Learn $P(f|e)$ from a parallel corpus
- Not sufficient data to estimate $P(f|e)$ directly

Statistical Modeling (2)



- Break the process into smaller steps

Statistical Modeling (3)

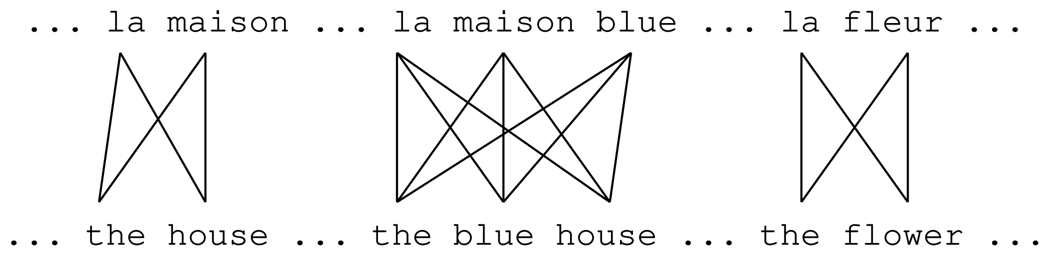


- Probabilities for smaller steps can be learned

Statistical Modeling (4)

- Generate a story how an English string e gets to be a foreign string f
  - Choices in story are decided by reference to parameters
    – e.g., $p(\text{bruja}|\text{witch})$
- Formula for $P(f|e)$ in terms of parameters
  - usually long and hairy, but mechanical to extract from the story
- Training to obtain parameter estimates from possibly incomplete data
  - off-the-shelf EM

Parallel Corpora

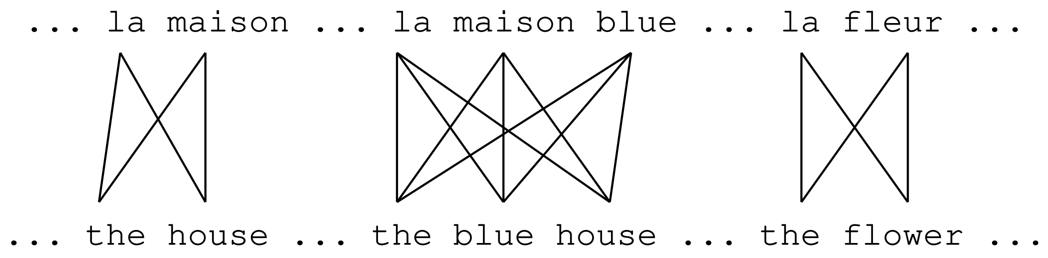


- Incomplete Data
  - English and foreign words, but no connections between them
- Chicken and egg problem
  - if we had the connections, we could estimate the parameters of our generative story
  - if we had the parameters, we could estimate the connections

EM Algorithm

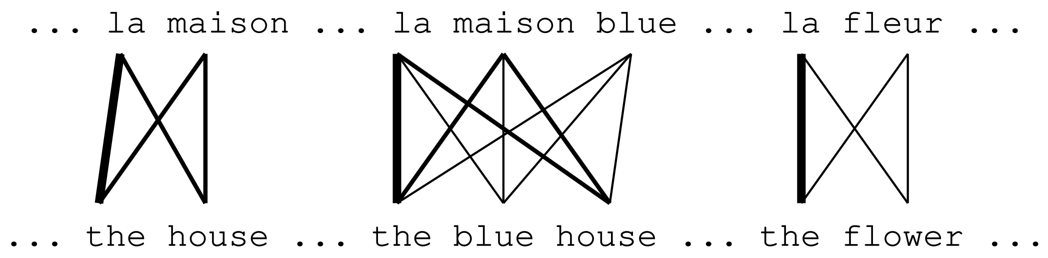
- Incomplete data
  - if we had complete data, would could estimate model
  - if we had model, we could fill in the gaps in the data
- EM in a nutshell
  - initialize model parameters (e.g. uniform)
  - assign probabilities to the missing data
  - estimate model parameters from completed data
  - iterate

EM Algorithm (2)



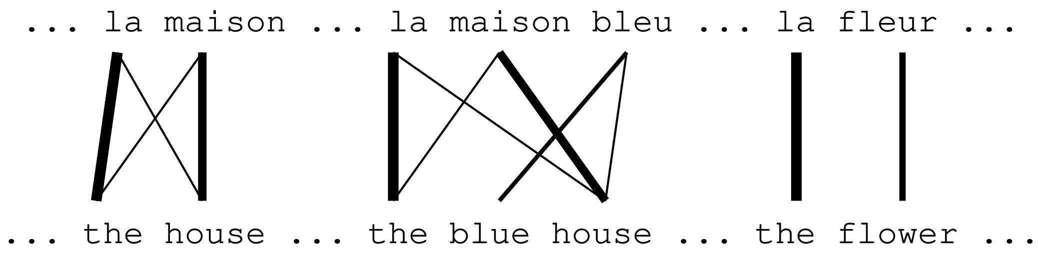
- Initial step: all connections equally likely
- Model learns that, e.g., la is often connected with the

EM Algorithm (3)



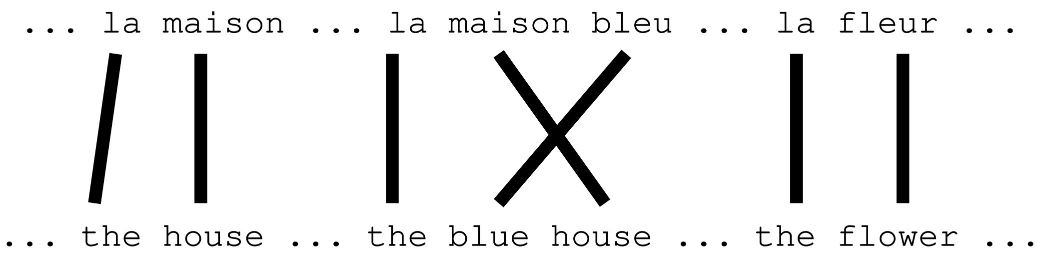
- After one iteration
- Connections, e.g., between la and the are more likely

EM Algorithm (4)



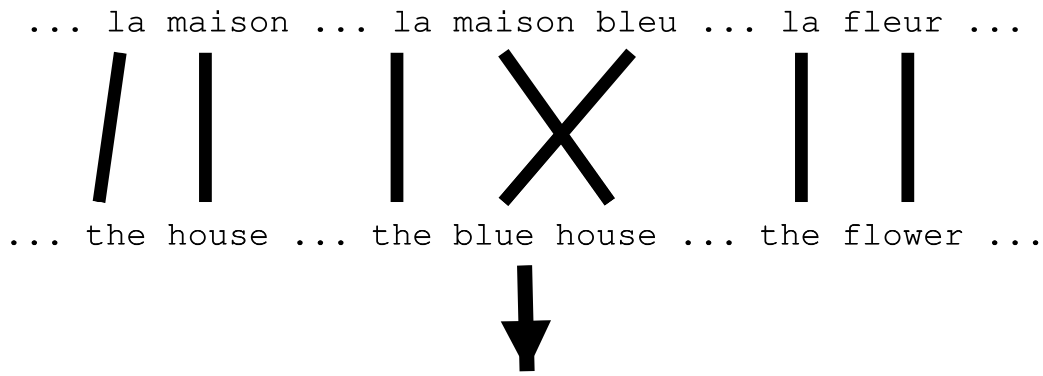
- After another iteration
- It becomes apparent that connections, e.g., between fleur and flower are more likely (pigeon hole principle)

EM Algorithm (5)



- Convergence
- Inherent hidden structure revealed by EM

EM Algorithm (6)




```
p(la|the) = 0.453
p(le|the) = 0.334
p(maison|house) = 0.876
p(bleu|blue) = 0.563
...
```

- Parameter estimation from the connected corpus

More detail on the IBM Models

- “A Statistical MT Tutorial Workbook” (Knight, 1999)
- “The Mathematics of Statistical Machine Translation” (Brown et al., 1993)
- Downloadable software: Giza++, ReWrite decoder

Outline

- Introduction to Statistical Machine Translation
- Translation Model: IBM Models and the EM Algorithm
- **Language Model**
- Decoding Algorithm
- Phrase-Based Machine Translation
- Log-Linear Models
- The Challenge of Syntax
- Divide and Conquer

Language Model

- Goal of the language model:
  detect good English

Language Model

- What is good English?
- Standard technique: trigram model
  - multiplication of trigram probabilities
  - p(witch|the green) > p(green|the witch)

```
Mary did not slap the green witch
Mary  =>  p(Mary)
Mary did   =>  p(did|Mary)
Mary did not  =>  p(not|Mary did)
     did not slap  =>  p(slap|did not)
         not slap the  =>  p(the|not slap)
             slap the green  =>  p(green|slap the)
                  the green witch  =>  p(witch|the green)
```


Using Web n-Grams as LM

- n-Grams seen on web:

	Human translation	**Machine translation**
bigrams	99% seen on web	97%
trigrams	97%	92%
4-grams	85%	80%
5-grams	65%	56%
6-grams	44%	32%
7-grams	30%	14%

- Successfully used web n-grams as feature [Koehn and Knight, 2003]

Exploiting Non-Parallel Corpora

- Use frequencies on the web [Soricut et al., 2002]
  - She has a lot of nerve. (20 Altavista)
  - It has a lot of nerve. (3 Altavista)
- Build suffix trees [Munteanu and Marcu, 2002]

Other Ideas in Language Modeling

- Multiple language models
- Language models on Part-of-Speech
- Syntactic language models (later more on this)
- Lots of LM ideas in speech recognition, many have not been tried yet in machine translation

Outline

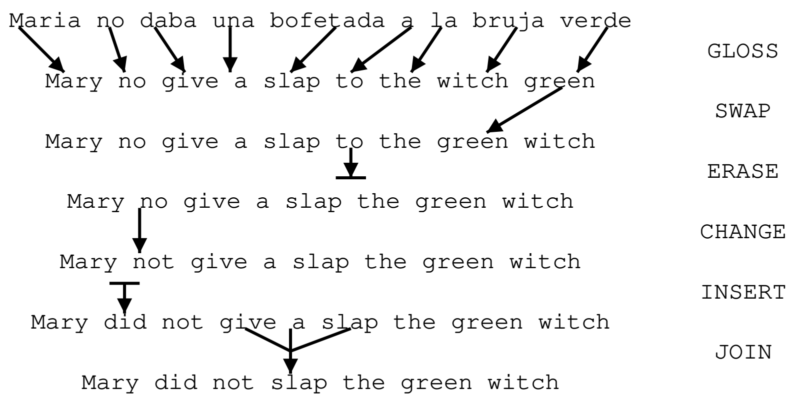
- Introduction to Statistical Machine Translation
- Translation Model: IBM Models and the EM Algorithm
- Language Model
- **Decoding Algorithm**
- Phrase-Based Machine Translation
- Log-Linear Models
- The Challenge of Syntax
- Divide and Conquer

Decoding Algorithm

- Goal of the decoding algorithm:

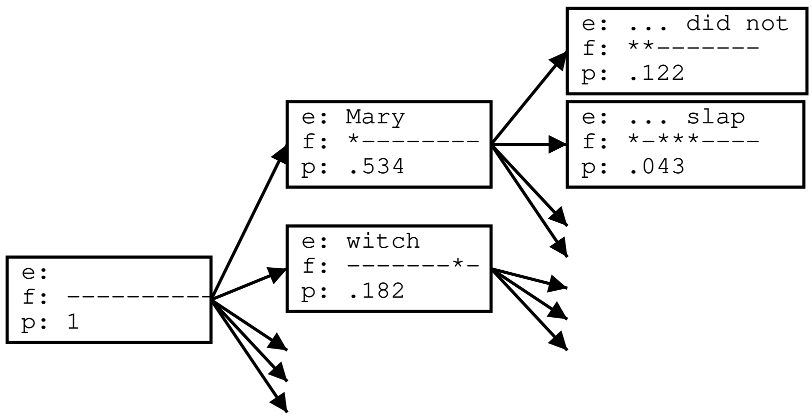
  Put models to work, perform the actual translation

Greedy Decoder



- Greedy hill climbing [Germann, 2003]
  - start with gloss
  - improve probability with actions
  - use 2-step look-ahead to avoid some local minima

Beam Search Decoding

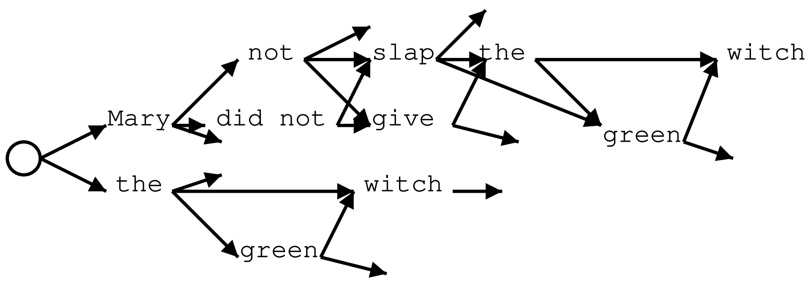


- Build English by hypothesis expansion
  - from left to right
  - search space exponential with sentence length
  - ⇒ reduction by pruning weak hypothesis

Beam: Search Space Reduction

- Organize hypotheses into bins
  - same foreign words covered (still exponential)
  - same number of foreign words covered
  - same number of English words generated
- Prune out weakest hypotheses in each bin
  - by absolute threshold (keep 100 best)
  - by relative cutoff (only if $<$0.01 worse than best)
- Future cost estimation
  - to have a more realistic comparison of hypothesis
  - compute expected cost of untranslated words
  - add to accumulated cost so far

Beam: Word Graphs



- Word graphs
  - search graph from beam search can be easily converted
  - important: hypothesis recombination
  - can be mined for n-best lists [Ueffing et al., 2002]

Other Decoding Methods

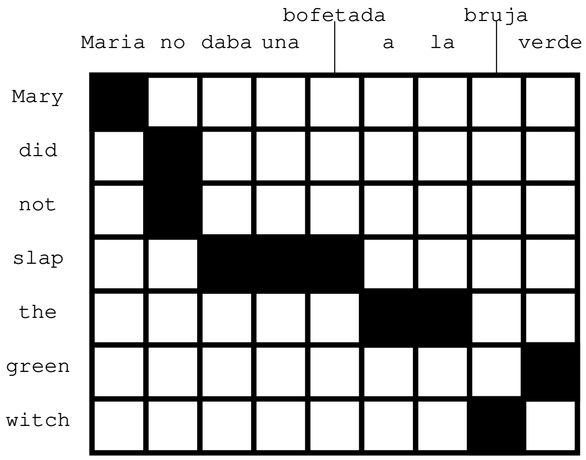
- Finite state transducers
  - e.g., [Al-Onaizan and Knight, 1998], [Alshawi et al., 1997]
  - well studied framework, many tools available
- Integer programming [Germann et al., 2001]

Outline

- Introduction to Statistical Machine Translation
- Translation Model: IBM Models and the EM Algorithm
- Language Model
- Decoding Algorithm
- **Phrase-Based Machine Translation**
- Log-Linear Models
- The Challenge of Syntax
- Divide and Conquer

Word Alignment

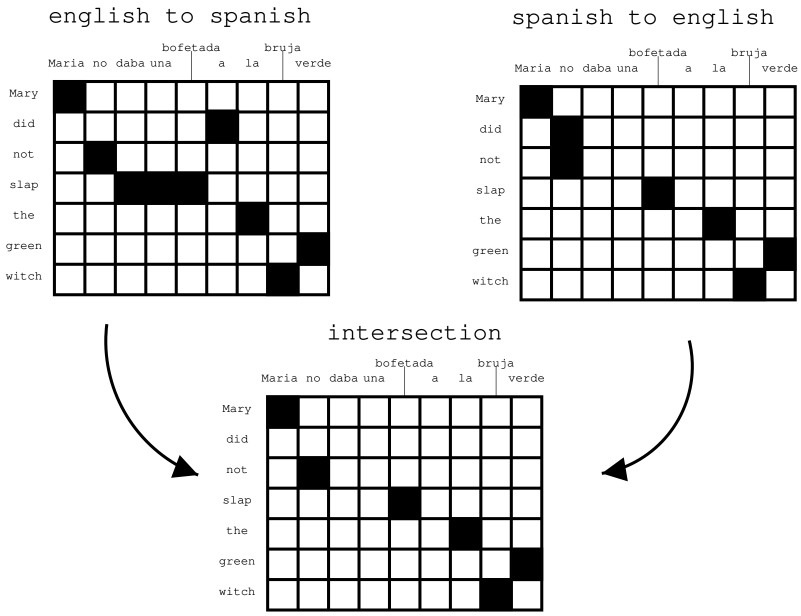
- Notion of word alignments valuable
- Trained humans can achieve high agreement
- Shared task at data-driven MT workshop at NAACL/HLT



Improved Word Alignments

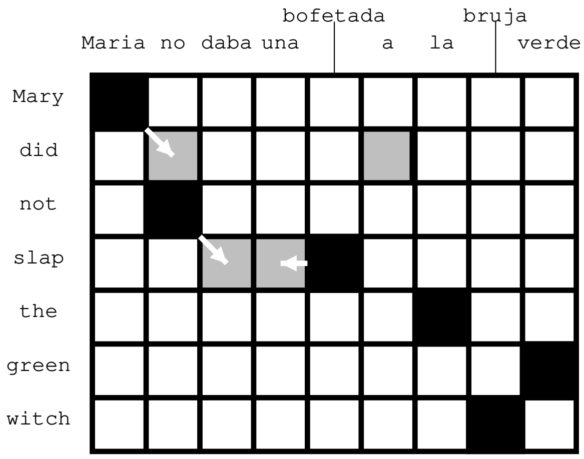
- Improving IBM Model word alignments with heuristics [Och and Ney, 2000, Koehn et al., 2003]
  - one-to-many problem of IBM Models
  - bidirectionally aligned corpora $e \to f$, $f \to e$
  - take intersection of alignment points (high precision, low recall)
  - grow additional alignment points (increase recall while preserving precision)

Improved Word Alignments (2)



- Intersection of bidirectional alignments

Improved Word Alignments (3)



- Grow additional alignment points

Improved Word Alignments (4)

- Heuristics for adding alignment points
  - only to directly neighboring
  - also to diagonally neighboring
  - also to non-neighboring
  - prefer English-foreign or foreign-to-English
  - use lexical probabilities or frequencies
  - extend only to unaligned words
  - ...

$\Rightarrow$ No clear advantage to any strategy
  - depends on corpus size
  - depends on language pair

Other Methods for Word Alignment

- Comparison of various methods in journal “Computation Linguistics”, March 2003 [Och and Ney, 2003]
- Shared task at data-driven MT workshop HLT/NAACL 2003

Flaws of Word-Based MT

- Multiple English words for one German word

German:	Zeitmangel	erschwert	das	Problem	.
Gloss:	LACK OF TIME	MAKES MORE DIFFICULT	THE	PROBLEM	.
Correct translation:	Lack of time makes the problem more difficult.				
MT output:	Time makes the problem .				

- Phrasal translation

German:	Eine	Diskussion	erübrigt	sich	demnach
Gloss:	A	DISCUSSION	IS MADE UNNECESSARY	ITSELF	THEREFORE
Correct translation:	Therefore, there is no point in a discussion.				
MT output:	A debate turned therefore .				

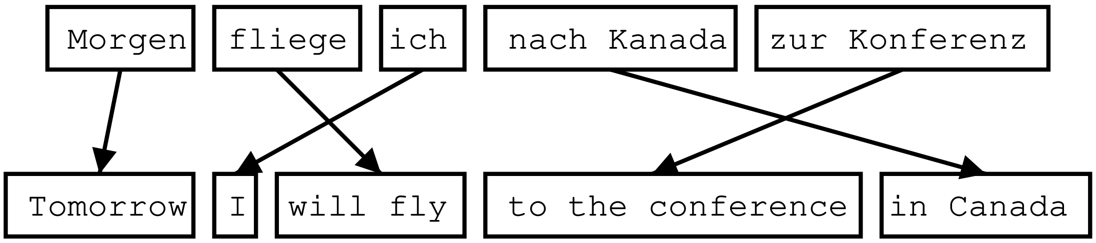
Flaws of Word-Based MT (2)

- Syntactic transformations

German:	Das	ist	der	Sache	nicht	angemessen	.
Gloss:	THAT	IS	THE	MATTER	NOT	APPROPRIATE	.
Correct translation:	That is not appropriate for this matter .						
MT output:	That is the thing is not appropriate .						

German:	Den	Vorschlag	lehnt	die	Kommission	ab	.
Gloss:	THE	PROPOSAL	REJECTS	THE	COMMISSION	OFF	.
Correct translation:	The commission rejects the proposal .						
MT output:	The proposal rejects the commission .						

Phrase-Based Translation



- Foreign input is segmented in phrases
  - any sequence of words, not necessarily linguistically motivated
- Each phrase is translated into English
- Phrases are reordered

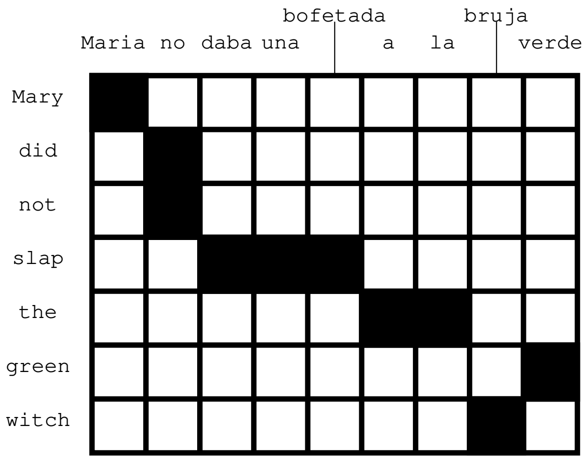
Advantages of Phrase-Based Translation

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

Three Phrase-Based Translation Models

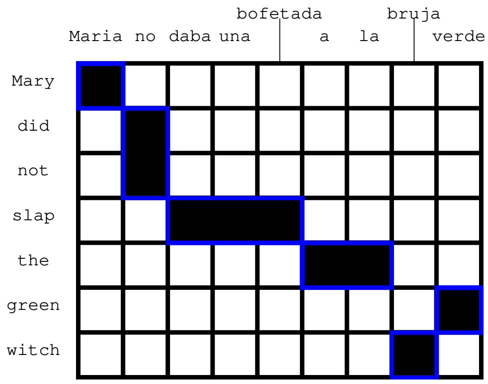
- Word alignment induced phrase model [Koehn et al., 2003]
- Alignment templates [Och et al., 1999]
- Joint phrase model [Marcu and Wong, 2002]

Word Alignment Induced Phrases



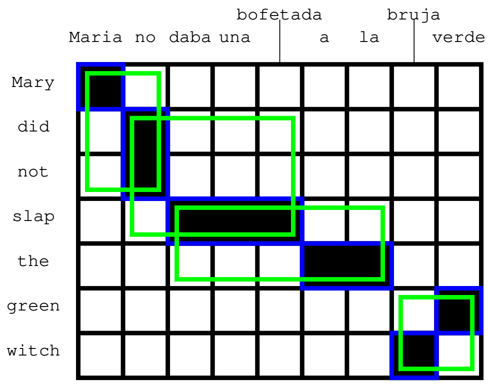
- Collect all phrase pairs that are consistent with the word alignment
  - a phrase alignment has to contain all alignment points for all words it covers

Word Alignment Induced Phrases (2)



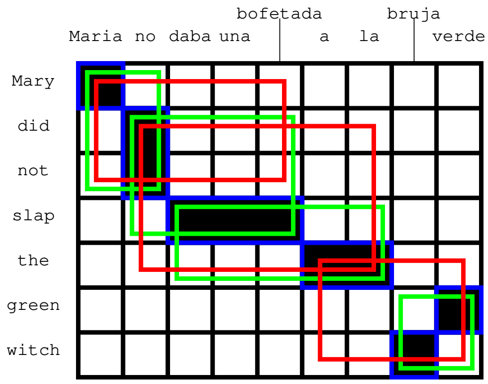
(Maria, Mary), (no, did not), (slap, daba una bofetada), (a la, the), (bruja, witch), (verde, green)

Word Alignment Induced Phrases (3)



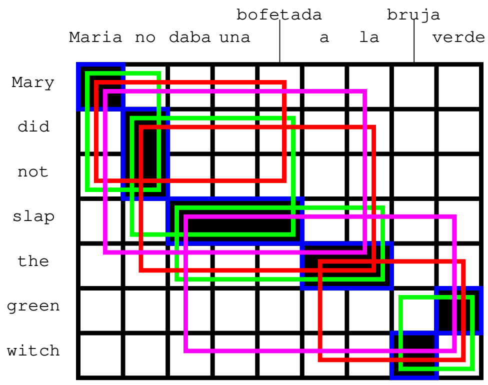
(Maria, Mary), (no, did not), (slap, daba una bofetada), (a la, the), (bruja, witch), (verde, green), (Maria no, Mary did not), (no daba una bofetada, did not slap), (daba una bofetada a la, slap the), (bruja verde, green witch)

Word Alignment Induced Phrases (4)



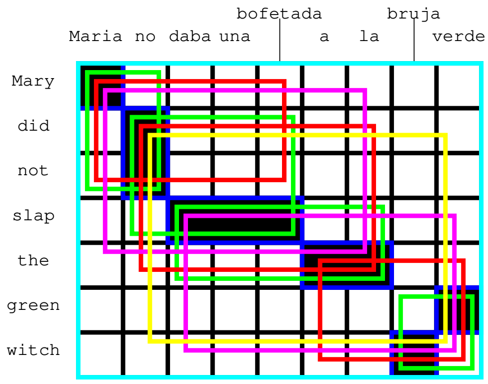
(Maria, Mary), (no, did not), (slap, daba una bofetada), (a la, the), (bruja, witch), (verde, green), (Maria no, Mary did not), (no daba una bofetada, did not slap), (daba una bofetada a la, slap the), (bruja verde, green witch), (Maria no daba una bofetada, Mary did not slap), (no daba una bofetada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

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)

Word Alignment Induced Phrases (6)



(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)

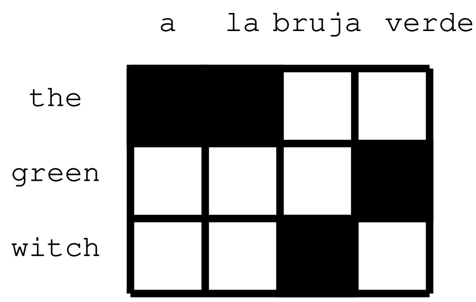
Word Alignment Induced Phrases (7)

- Given the collected phrase pairs, estimate the phrase translation probability distribution by relative frequency:

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

- No smoothing is performed

Word Alignment Induced Phrases (8)



- Lexical weighting:

$$p_w(\bar{f}|\bar{e}, a) = \prod_{i=1}^{n} \frac{1}{|\{j|(i,j) \in a\}|} \sum_{\forall (i,j) \in a} w(f_i|e_j)$$

p_w(a la bruja verde|the green witch) $=$

w(a|the) $\times$ w(la|the) $\times$

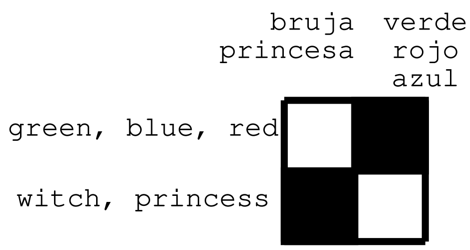
w(verde|green) $\times$

w(bruja|witch)

Other Phrase Extraction Schemes

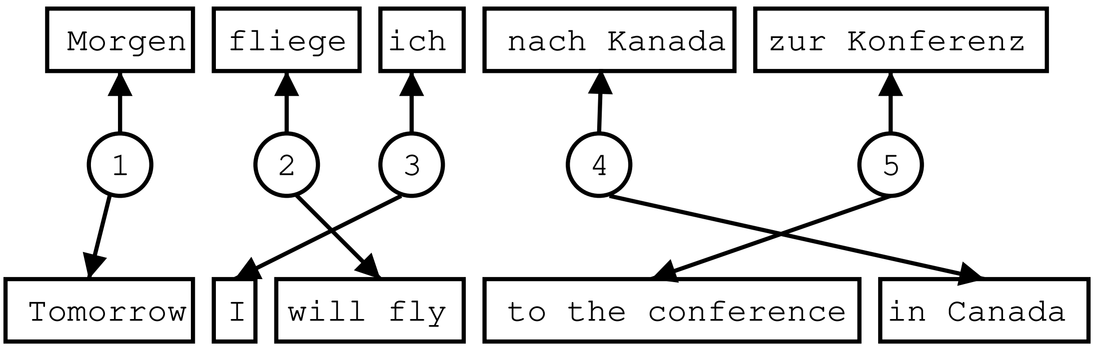
- Using sure and potential alignment points [Tillmann, ACL 2003]
- Allow violation of consistency with word alignment [Venugopal et al., ACL 2003]
- What works best? Unclear at this point

Alignment Templates [Och et al., 1999]



- Word classes instead of words
  - alignment templates instead of phrases
  - more reliable statistics for translation table
  - smaller translation table
  - more complex decoding
- Same lexical weighting

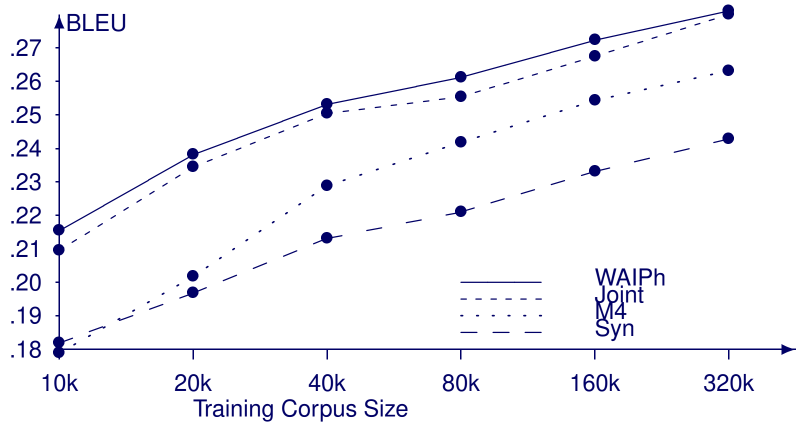
Joint Phrase Model



- Direct phrase alignment of parallel corpus [Marcu and Wong, 2002]
- Generative story
  - a number of concepts are created
  - each concept generates a foreign and English phrase
  - the English phrases are reordered

Evaluation of Phrase Models

- Direct comparison of models [Koehn et al., 2003]
  - results improve log-linear with training corpus size
  - WAIPh slightly better than Joint (same decoder, same LM)
  - better than IBM Model 4 (different decoder)
  - using only phrases that are syntactic constituents hurts



Evaluation of Phrase Models (2)

- Different language pairs
  - results for WAIPh
  - better than IBM Model 4
  - lexical weighting always helps

Language Pair	Model4	Phrase	Lex
English-German	0.20	0.24	0.24
French-English	0.28	0.33	0.34
English-French	0.26	0.31	0.32
Finnish-English	0.22	0.27	0.28
Swedish-English	0.31	0.35	0.36
Chinese-English	0.12	0.14	0.14

Limits of Phrase Models

- Non-contiguous phrases
  - German: *Ich habe das Auto gekauft*
  - English: *I bought the car*
  - good phrase pair: *habe ... gekauft == bought*
- Syntactic transformations
  - German: *Den Antrag verabschiedet das Parlament*
  - English gloss: *The draft approves the Parliament*
  - case marking that indicates that "the draft" is object is lost during translation

Phrase-Based MT: Do it yourself

- Phrase-based MT has currently best performance
- Corpora available at LDC, ISI, other places
  - e.g., Europarl `http://www.isi.edu/~koehn/europarl/`
- Giza++ toolkit available at RWTH Aachen

  `http://www-i6.informatik.rwth-aachen.de/web/Software/GIZA++.html`
- Language model available at SRI
  - `http://www.speech.sri.com/projects/srilm/`
- Pharaoh decoder available at ISI
  - contact `koehn@isi.edu`

Outline

- Introduction to Statistical Machine Translation
- Translation Model: IBM Models and the EM Algorithm
- Language Model
- Decoding Algorithm
- Phrase-Based Machine Translation
- **Log-Linear Models**
- The Challenge of Syntax
- Divide and Conquer

Log-Linear Models

- Many different knowledge sources useful
  - language model
  - reordering (distortion) model
  - phrase translation Model
  - word translation model
  - word penalty
  - additional language models
  - additional features

Log-Linear Models (2)

- IBM Models provided mathematical justification for factoring components together
  - $p_{LM} \times p_{TM} \times p_D$
- More components p_i, weighting λ_i

$\Rightarrow \prod_i p_i^{\lambda_i} =$

= $exp(\sum_i \lambda_i log(p_i))$

Learning the Weights

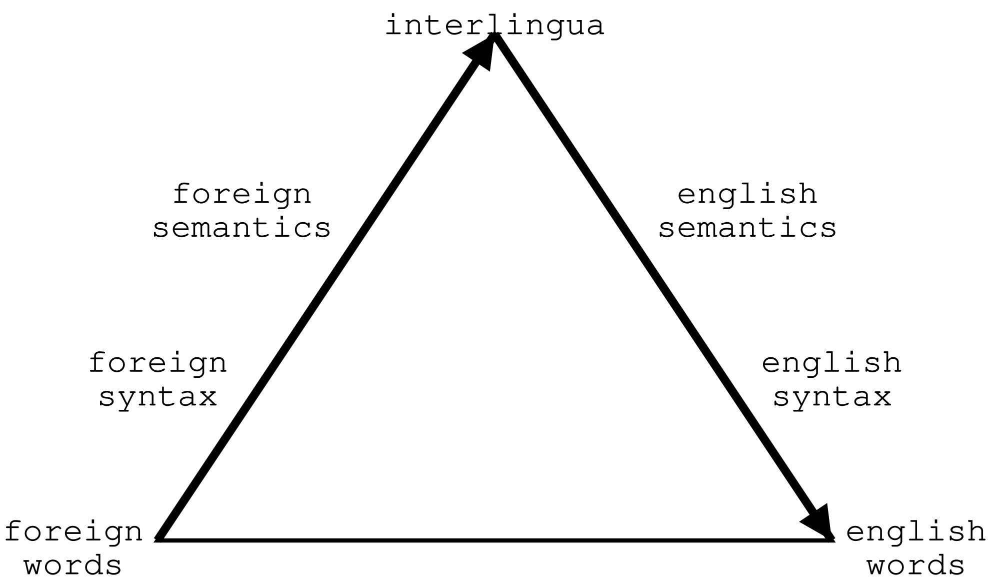
- Contribution of components p_i determined by weight λ_i

$\Rightarrow$ Learn these weights from small held-out (“development”) corpus

  - set the weights, so that optimal translation performance on this development corpus is achieved
  - requires automatic scoring method (e.g., BLEU)
- Methods
  - manual picking of weights
  - greedy search
  - discriminative maximum entropy training [Och, 2002]
  - minimum error rate training [Och, 2003]

Outline

The Challenge of Syntax

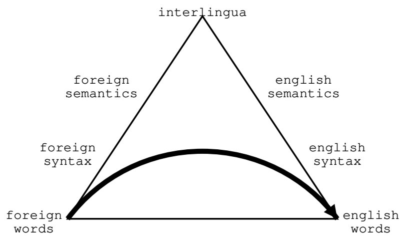


- Remember the pyramid

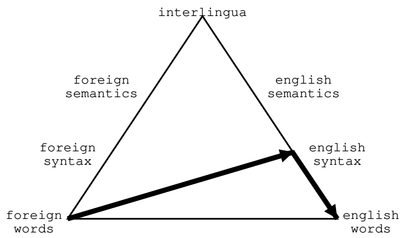
Advantages of Syntax-Based Translation

- Reordering for syntactic reasons
  - e.g., move German object to end of sentence
- Better explanation for function words
  - e.g., prepositions, determiners
- Conditioning to syntactically related words
  - translation of verb may depend on subject or object
- Use of syntactic language models

Syntax-Based Translation Models



Wu [1997], Alshawi et al. [1998]



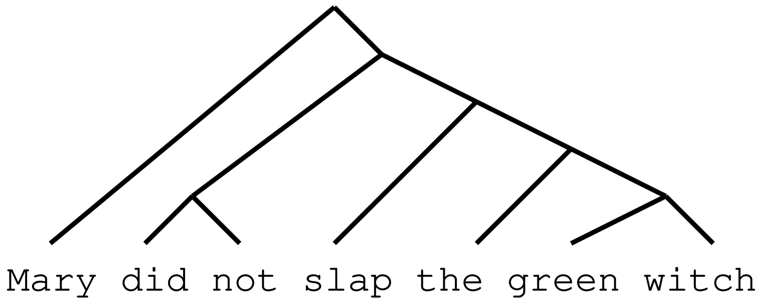
Yamada and Knight [2001]

Inversion Transduction Grammars

- Generation of both English and foreign trees [Wu, 1997]
- Rules (binary and unary)
  - $A \rightarrow A_1 A_2 | A_1 A_2$
  - $A \rightarrow A_1 A_2 | A_2 A_1$
  - $A \rightarrow e | f$
  - $A \rightarrow e | *$
  - $A \rightarrow * | f$

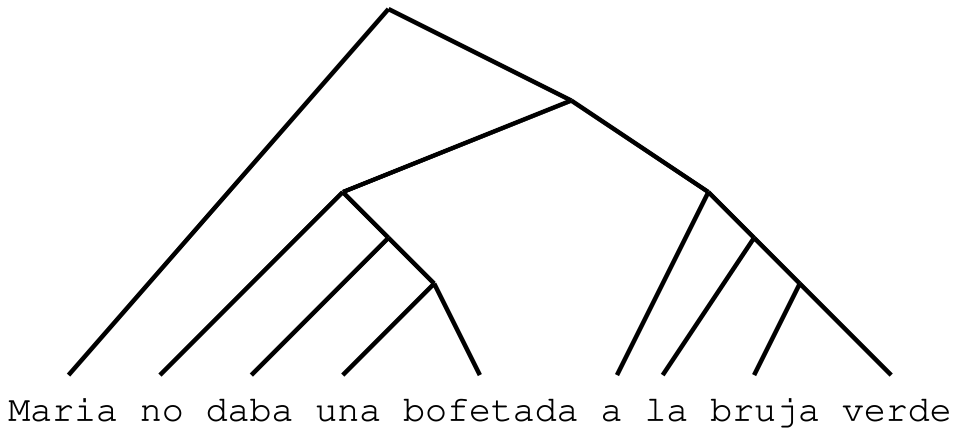
$\Rightarrow$ Common binary tree required
  - limits the complexity of reorderings

Syntax Trees



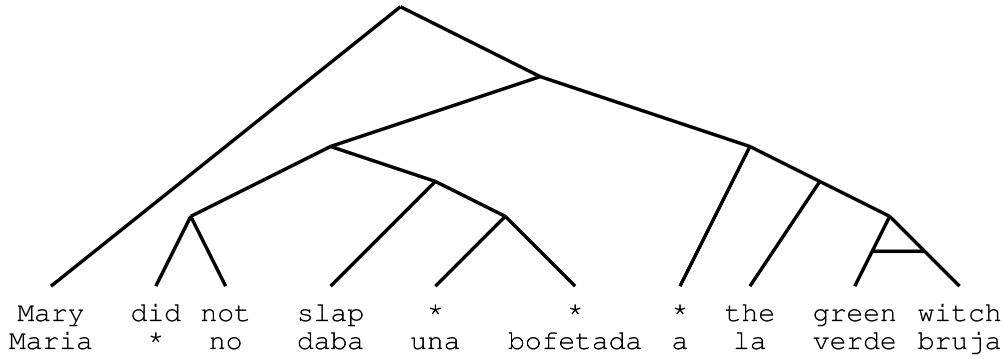
- English binary tree

Syntax Trees (2)



- Spanish binary tree

Syntax Trees (3)

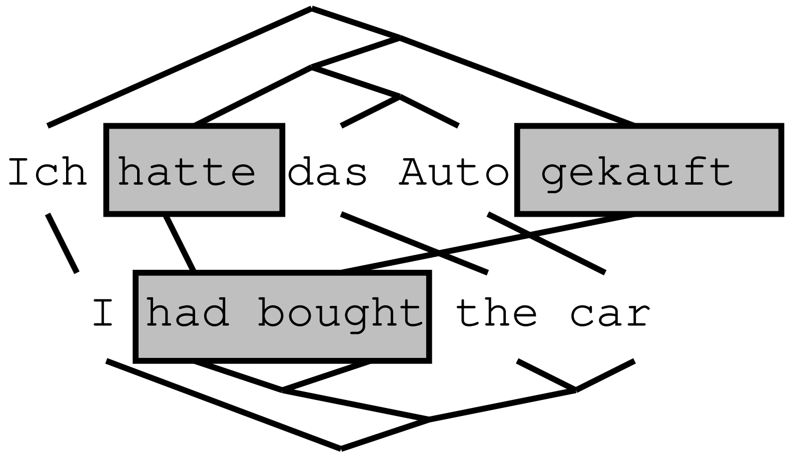


- Combined tree with reordering of Spanish

Hierarchical Transduction Models

- Based on finite state transducers [Alshawi et al., 1998]
  - also common binary tree required
  - lexicalized non-terminal rules
- Generation of sentence pair
  1. create initial **head word** (e.g., [daba : slap])
  2. extend head word by **adding dependents** (e.g., [bruja : witch]); foreign and English could be placed on different sides of head; dependents could be single word, empty, or phrases
  3. pick one of the dependents as new head word for extension (step 2); or terminate

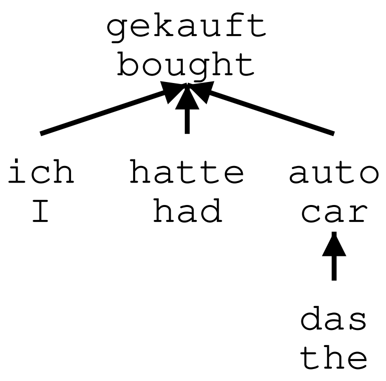
Common Binary Tree Requirement



- No common binary tree possible
- Maybe most languages are syntactically too different?

⇒ Jump ahead to semantics

Dependency Structure

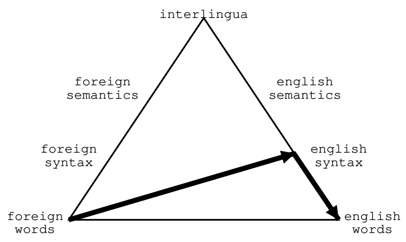


- Common dependency tree
- Interest in dependency-based translation models
  - e.g. Czech-English [Cmejrek et al., 2003]
  - current systems mixed statistical/rule-based
  - probably good generation system necessary

Direct Correspondence Assumption

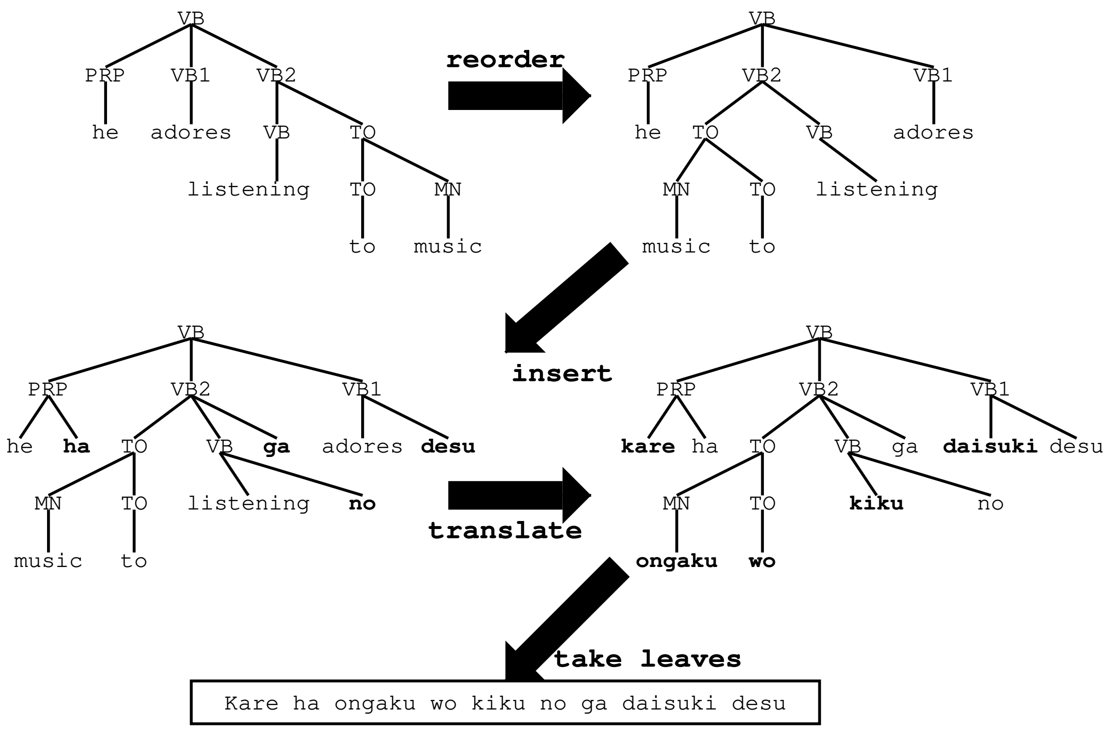
- Do foreign and English have same dependency structure?
- Direct correspondence assumption [Hwa et al., 2002]
  - empirical study (by projection) of Chinese-English parallel corpus
  - even with modifications, only 67% precision/recall
  - more structure could be preserved, if tried

String to Tree Translation



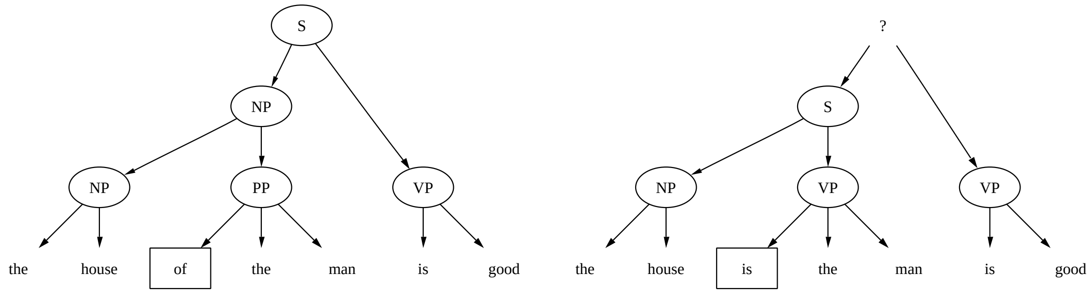
- Use of English syntax trees [Yamada and Knight, 2001]
  - exploit rich resources on the English side
  - obtained with statistical parser [Collins, 1997]
  - flattened tree to allow more reorderings
  - works well with syntactic language model

Yamada and Knight [2001]



Syntactic Language Model

- Good syntax tree → good English
- Allows for long distance constraints



- Left translation preferred by syntactic LM

String to Tree Transfer and Syntactic LM

- Work presented at this MT Summit by Charniak, Knight, Yamada
  - more grammatical correct output
  - more perfectly translated sentences
  - ... but no improvement in BLEU
- Syntactic transfer and LM on top of phrase translation
  - parse a lattice generated by phrase-based MT
  - no results yet

Crossings

- Do English trees match foreign strings?
- Crossings between French-English [Fox, 2002]
  - 0.29-6.27 per sentence, depending on how it is measured
- Can be reduced by
  - flattening tree, as done by [Yamada and Knight, 2001]
  - detecting phrasal translation
  - special treatment for small number of constructions
- Most coherence between dependency structures

Augment Models with Syntactic Features

- Intuition: other models work fine, syntax provides additional clues
- Define syntactic properties that should hold
  - preservation of plural
  - output should have verb
  - no dropping of content words
  - ...
- 2003 summer workshop at John Hopkins: little improvement

Full Syntactic/Semantic Translation

- Existing systems hybrid rule-based / statistical
  - Czech-English [Cmejrek et al., 2003]
  - Spanish-English [Habash, 2002]
- Performance below phrase-based statistical systems
- Why is it so hard?
  - loss of good phrasal translations [Koehn et al., 2003]
  - lack of foreign syntactic parsers
  - differences in syntactic structure
  - semantic transfer hard to learn (no parallel data)

Outline

- Introduction to Statistical Machine Translation
- Translation Model: IBM Models and the EM Algorithm
- Language Model
- Decoding Algorithm
- Phrase-Based Machine Translation
- Log-Linear Models
- The Challenge of Syntax
- **Divide and Conquer**

Divide and Conquer

- Named Entities
  - names
  - numbers
  - dates
  - quantities
- Noun Phrases

Numbers, Dates, Entities

- Translation tables for numbers?

f	e	p(f|e)
2003	2003	0.7432
2003	2000	0.0421
2003	year	0.0212
2003	the	0.0175
2003	...	...

- Or by special handling?
  - XML markup of MT input [Germann et al., 2003]
  - `the revenue for`
    `<number translate-as=''2003''> 2003 < /number>`
    `is higher than ...`
  - same for dates and quantities
  - infinite variety, but simple translation rules

Names

- Often not in training corpus
- Require special treatment
- Issues
  - recognition of name vs. non-name
  - translation (Defense Department) vs. transliteration (George Bush)
  - especially hard, if different character set (Arabic, Chinese, Cyrillic, ...)
- Phonetic reasoning and web resources

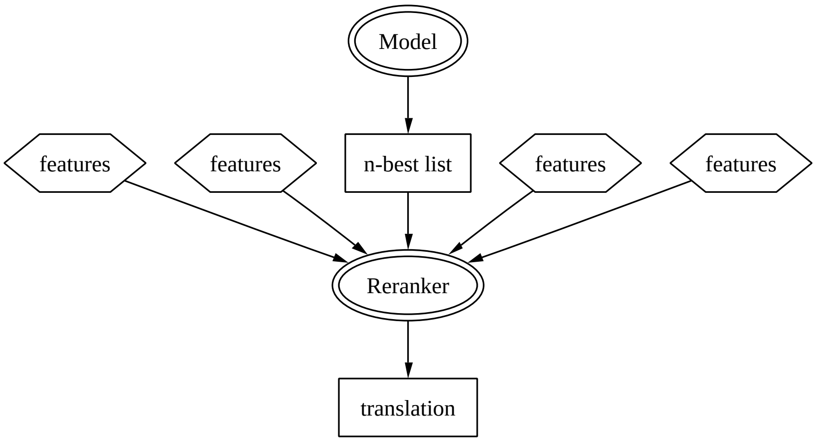
Arabic-English	**all**	person	organization	location
Sakhr	**61%**	47%	81%	36%
[Al-Onaizan and Knight, 2002]	**73%**	64%	87%	51%
Human	**75%**	68%	95%	42%

Noun Phrases

- Noun Phrases can be translated in separation [Koehn and Knight, 2003]
  - German-English: 75% are, 98% can be
  - also other examined languages: Portuguese-E, Chinese-E
- Definition of NP/PP
  - (informally): maximal phrases that contain at least one noun and no verb
  - *(The permanent tribunal)* is designed to prosecute *(individuals)* *(for genocide, crimes against humanity and other war crimes)* .
  - cover about half of the words, all nouns (largest open word class)
  - shorter, simpler than full sentences

$\Rightarrow$ special linguistic modeling, expensive features

Noun Phrases: Re-Ranking

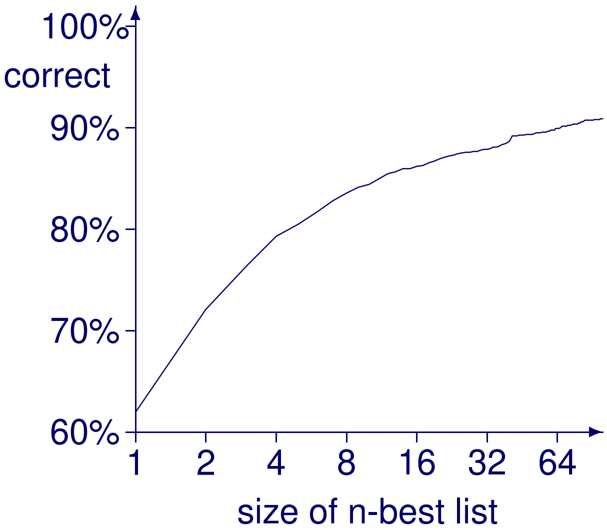


- Maximum entropy reranking
  - allows for variety of features: binary, integer, real-valued
  - see also direct maximum entropy models [Och and Ney, 2002]

Noun Phrases: Re-Ranking (2)

- Correct translations in the n-best list

  → over 90% accuracy possible with 100-best list reranking



Noun Phrases: Results

- Results for German-English

System	NP/PP Correct	BLEU Full Sentence
IBM Model 4	53.2%	0.172
Phrase Model	58.7%	0.188
Compound Splitting	61.5%	0.195
Re-Estimated Parameters	63.0%	0.197
Web Count Features	64.7%	0.198
Syntactic Features	65.5%	0.199

How Good is Statistical MT?

- Out-of-domain (Sports)

Basketball Network and Valve Promoted More Eastern Second Round
Washington (Afp) new Jersey nets basketball team Thursday again rather than Indian it slipped horseback birds will be Miller of selling your life and hard work, the two extensions to competition after more than 120 109 to Clinton slipped horseback, winning more quarter after the competition for the first round matches of the war, and promoted the second round...

- In-domain (Politics)

The United States and India May Will Be Held in the Past 40 Years the First Joint Military Exercises
(Afp report from new Delhi) India and U. S. will be held in the past 39 years the first joint military exercises in the world's two biggest democracies the cooperative relationship between making milestone.
The Defense Ministry said in a class Indian paratrooper Brigade mid-May and the US Pacific Command of the special units in the well-known far and near the Thai women Maha tomb near joint military exercises.
The two countries will provide air support.

- DARPA Chinese-English task (fairly hard)
  - This is actual output of the ISI system