

Data Intensive Linguistics — Lecture 16

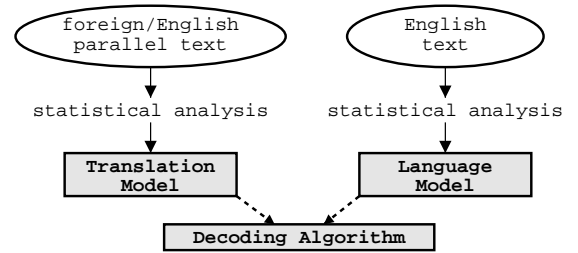
Machine translation (III): Decoding

Philipp Koehn
6 March 2006

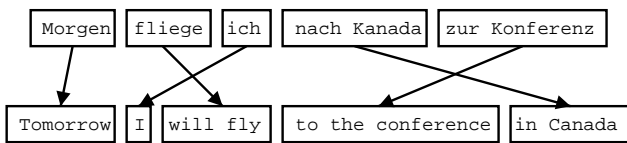


Statistical Machine Translation

- Components: Translation model, language model, decoder



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

Phrase Translation Table

- Phrase Translations for "den Vorschlag":

English	$\phi(e f)$	English	$\phi(e f)$
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

Decoding Process

Maria	no	dio	una	bofetada	a	la	bruja	verde
-------	----	-----	-----	----------	---	----	-------	-------

- Build translation left to right
 - select foreign words to be translated

Decoding Process

Maria	no	dio	una	bofetada	a	la	bruja	verde
-------	----	-----	-----	----------	---	----	-------	-------

- Build translation *left to right*
 - select foreign words to be translated
 - find English phrase translation
 - add English phrase to end of partial translation

Decoding Process

Maria	no	dio	una	bofetada	a	la	bruja	verde
-------	----	-----	-----	----------	---	----	-------	-------

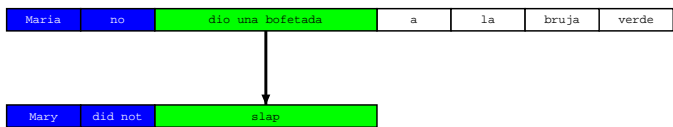
- Build translation left to right
 - select foreign words to be translated
 - find English phrase translation
 - add English phrase to end of partial translation
 - mark foreign words as translated

Decoding Process

Maria	no	dio	una	bofetada	a	la	bruja	verde
-------	----	-----	-----	----------	---	----	-------	-------

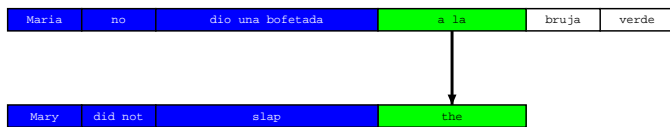
- One to many translation

Decoding Process



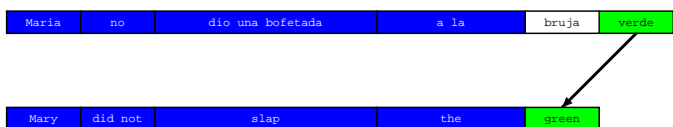
- Many to one translation

Decoding Process



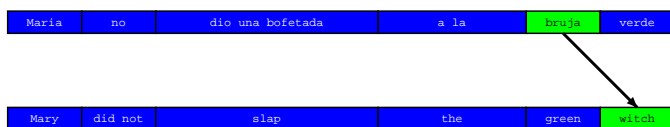
- *Many to one* translation

Decoding Process



- *Reordering*

Decoding Process



- Translation *finished*

Translation Options

Maria	no	dio	una	bofetada	a	la	bruja	verde
Mary	not	give	a	slap	to	the	witch	green
	did not		a	slap	by		green	witch
	no		slap		to	the		
	did not give				to			
					the			
				slap		the	witch	

- Look up *possible phrase translations*
 - many different ways to *segment* words into phrases
 - many different ways to *translate* each phrase

Hypothesis Expansion

Maria	no	dio	una	bofetada	a	la	bruja	verde
Mary	not	give	a	slap	to	the	witch	green
	did not		a	slap	by		green	witch
	no		slap		to	the		
	did not give				to			
					the			
				slap		the	witch	

e: Mary
f: 0.534
p: 1

- Start with *empty hypothesis*
 - e: no English words
 - f: no foreign words covered
 - p: probability 1

Hypothesis Expansion

Maria	no	dio	una	bofetada	a	la	bruja	verde
Mary	not	give	a	slap	to	the	witch	green
	did not		a	slap	by		green	witch
	no		slap		to	the		
	did not give				to			
					the			
				slap		the	witch	

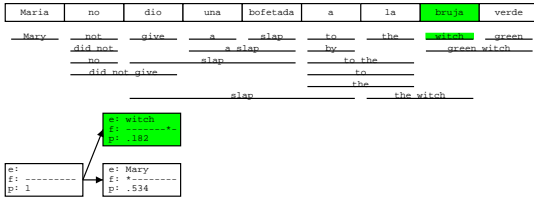
e: Mary
f: 0.534
p: 1

- Pick *translation option*
- Create *hypothesis*
 - e: add English phrase Mary
 - f: first foreign word covered
 - p: probability 0.534

A Quick Word on Probabilities

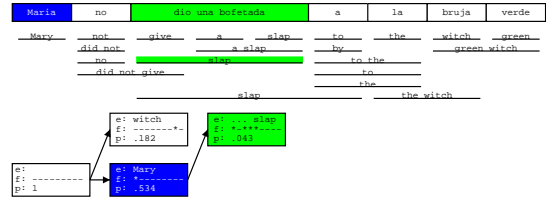
- Not going into detail here, but...
- *Translation Model*
 - phrase translation probability $p(\text{Mary}|\text{Maria})$
 - reordering costs
 - phrase/word count costs
 - ...
- *Language Model*
 - uses trigrams:
 - $p(\text{Mary did not}) = p(\text{Mary}|\text{START}) \times p(\text{did}|\text{Mary,START}) \times p(\text{not}|\text{Mary did})$

Hypothesis Expansion



- Add another hypothesis

Hypothesis Expansion



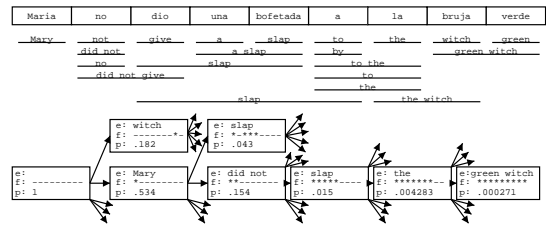
- Further hypothesis expansion

Hypothesis Expansion



- ... until all foreign words covered
 - find best hypothesis that covers all foreign words
 - backtrack to read off translation

Hypothesis Expansion



- Adding more hypothesis
- ⇒ Explosion of search space

Explosion of Search Space

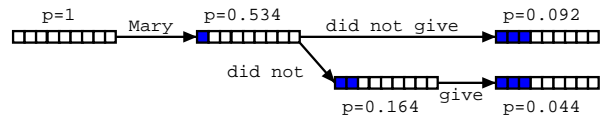
- Number of hypotheses is exponential with respect to sentence length

⇒ Decoding is NP-complete [Knight, 1999]

⇒ Need to reduce search space

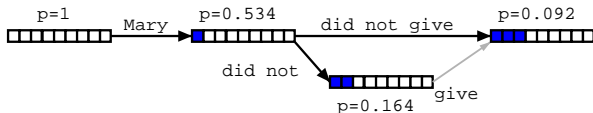
- risk free: hypothesis recombination
- risky: histogram/threshold pruning

Hypothesis Recombination



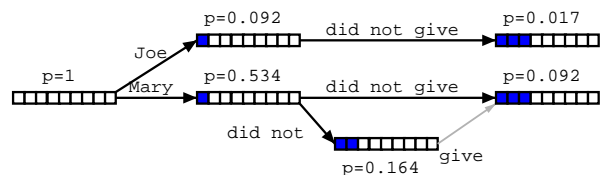
- Different paths to the same partial translation

Hypothesis Recombination



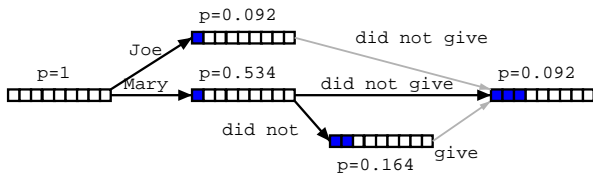
- Different paths to the same partial translation
- ⇒ Combine paths
- drop weaker path
 - keep pointer from weaker path (for lattice generation)

Hypothesis Recombination



- Recombined hypotheses do not have to match completely
- No matter what is added, weaker path can be dropped, if:
 - last two English words match (matters for language model)
 - foreign word coverage vectors match (effects future path)

Hypothesis Recombination



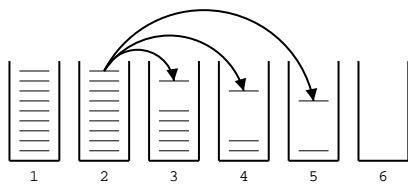
- Recombined hypotheses do not have to match completely
- No matter what is added, weaker path can be dropped, if:
 - last two English words match (matters for language model)
 - foreign word coverage vectors match (effects future path)

⇒ *Combine paths*

Pruning

- Hypothesis recombination is *not sufficient*
- ⇒ Heuristically *discard* weak hypotheses early
- Organize Hypothesis in **stacks**, e.g. by
 - same foreign words covered
 - same number of foreign words covered
 - same number of English words produced
- Compare hypotheses in stacks, discard bad ones
 - histogram pruning**: keep top n hypotheses in each stack (e.g., $n=100$)
 - threshold pruning**: keep hypotheses that are at most α times the cost of best hypothesis in stack (e.g., $\alpha = 0.001$)

Hypothesis Stacks



- Organization of hypothesis into stacks
 - here: based on *number of foreign words* translated
 - during translation all hypotheses from one stack are expanded
 - expanded Hypotheses are placed into stacks

Comparing Hypotheses

Comparing hypotheses with *same number of foreign words* covered

Maria no dio una bofetada a la bruja verde

e: Mary did not
f: **-----
p: 0.154

better
partial
translation

e: the
f: -----+*--
p: 0.354

covers
easier part
--> lower cost

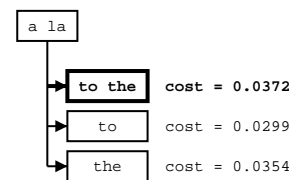
- Hypothesis that covers *easy part* of sentence is preferred
- ⇒ Need to consider **future cost** of uncovered parts

Future Cost Estimation



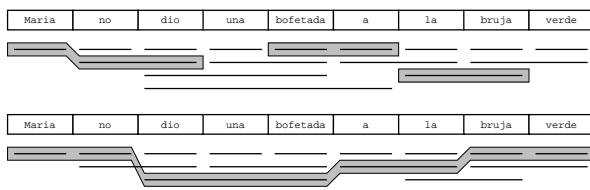
- Estimate cost* to translate remaining part of input
 - Step 1: estimate future cost for each *translation option*
 - look up translation model cost
 - estimate language model cost (no prior context)
 - ignore reordering model cost
- $LM * TM = p(\text{to}) * p(\text{the}|\text{to}) * p(\text{to the}|\text{a la})$

Future Cost Estimation: Step 2



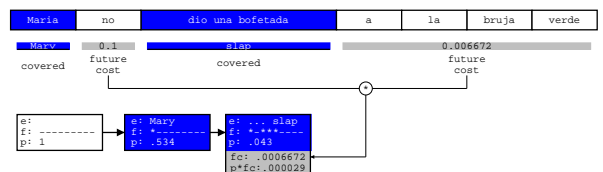
- Step 2: find *cheapest cost* among translation options

Future Cost Estimation: Step 3



- Step 3: find *cheapest future cost path* for each span
 - can be done *efficiently* by dynamic programming
 - future cost for every span can be *pre-computed*

Future Cost Estimation: Application



- Use future cost estimates when *pruning* hypotheses
- For each *uncovered contiguous span*:
 - look up *future costs* for each maximal contiguous uncovered span
 - add* to actually accumulated cost for translation option for pruning

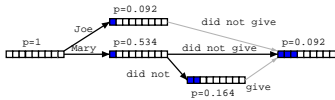
A* search

- Pruning might drop hypothesis that lead to the best path ([search error](#))
- **A* search**: safe pruning
 - future cost estimates have to be accurate or underestimates
 - **lower bound** for probability is established early by **depth first search**: compute cost for one complete translation
 - if cost-so-far and future cost are worse than **lower bound**, hypothesis can be safely discarded
- Not commonly done, since not aggressive enough

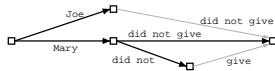
Limits on Reordering

- Reordering may be **limited**
 - **Monotone** Translation: No reordering at all
 - Only phrase movements of at most n words
- Reordering limits **speed** up search (polynomial instead of exponential)
- Current reordering models are weak, so limits **improve** translation quality

Word Lattice Generation



- **Search graph** can be easily converted into a **word lattice**
 - can be further mined for **n-best lists**
 - enables **reranking** approaches
 - enables **discriminative training**



Sample N-Best List

- **Simple N-best list**:

Translation	Reordering	LM	TM	WordPenalty	Score
this is a small house	0	-27.0908	-1.83258	-5	-28.9234
this is a little house	0	-28.1791	-1.83258	-5	-30.0117
it is a small house	0	-27.108	-3.21888	-5	-30.3965
it is a little house	0	-28.1963	-3.21888	-5	-31.4152
this is an small house	0	-31.7294	-1.83258	-5	-33.562
it is an small house	0	-32.3094	-3.21888	-5	-35.5283
this is an little house	0	-33.7639	-1.83258	-5	-35.5965
this is a house small	-3	-31.4851	-1.83258	-5	-36.3176
this is a house little	-3	-31.5689	-1.83258	-5	-36.4015
it is an little house	0	-34.2439	-3.21888	-5	-37.5628
it is a house small	-3	-31.5022	-3.21888	-5	-37.7211
this is an house small	-3	-32.8999	-1.83258	-5	-37.7325
it is a house little	-3	-31.586	-3.21888	-5	-37.9049
this is an house little	-3	-32.9837	-1.83258	-5	-37.8163
the house is a little	-7	-28.5107	-2.52573	-5	-38.0364
the is a small house	0	-35.6899	-2.52573	-5	-38.2156
is it a little house	-4	-30.9603	-3.91202	-5	-38.2723
the house is a small	-7	-28.7683	-2.52573	-5	-38.294
it 's a small house	0	-34.8557	-3.91202	-5	-38.7677
this house is a little	-7	-28.0443	-3.91202	-5	-38.9563
it 's a little house	0	-35.1446	-3.91202	-5	-39.0566
this house is a small	-7	-28.3018	-3.91202	-5	-39.2139

XML Markup

Er erzielte <NUMBER english='17.55'>17,55</NUMBER> Punkte .

- **Add additional translation options**
 - number translation
 - name translation
- **Additional options**
 - provide multiple translations
 - provide probability distribution along with translations
 - allow bypassing of provided translations