

Decoding



Phrase-Based Decoding

- Inputs:
 - Language model that scores $P(e_i|e_1, \dots, e_{i-1}) \approx P(e_i|e_{i-n-1}, \dots, e_{i-1})$
 - Phrase table: set of phrase pairs (\mathbf{e}, \mathbf{f}) with probabilities $P(\mathbf{f}|\mathbf{e})$
- What we want to find: \mathbf{e} produced by a series of phrase-by-phrase translations from an input \mathbf{f}



Phrase lattices are big!

这	7人	中包括	来自	法国	和	俄罗斯	的	宇航	员	.
the	7 people	including	by some	and	the russian	the	the astronauts			,
it	7 people	included	by france	and the	the russian		international astronautical	of rapporteur .		,
this	7 out	including the	from	the french	and the russian	the fifth	space	members		.
these	7 among	including from		of the french	and of the russian	of	aerospace	members .		.
that	7 persons	including from	the	of france	and to russian	of the	astronauts	members .		.
	7 include		from the	of france and	russian					the
	7 numbers include		from france	and russian			of astronauts who			.
	7 populations include		those from france	and russian			astronauts .			.
	7 deportees included		come from	france	and russia	in	astronautical	personnel		;
	7 philtrum	including those from		france and	and russia	in a space	astronaut	member		.
		including representatives from		france and the	russia		astronaut			.
		include	came from	france and russia		by cosmonauts				.
		include representatives from	french	and russia		cosmonauts				.
		include	came from france	and russia's		cosmonaut .				.
		includes	coming from	french and	russia's	's	astronaut	member .		.
				french and russian	and russia	astronauts				.
				french	and russia's		special rapporteur			.
					, and russia		rapporteur			.
					, and russia		rapporteur .			.
					, and russia					.
				or	russia's					.

Slide credit: Dan Klein



Monotonic Translation

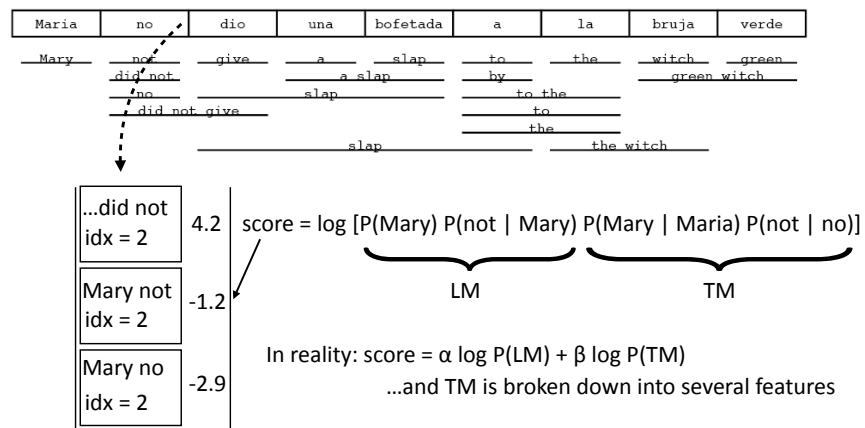
Maria	no	dio	una	bofetada	a	la	bruja	verde
Mary	not	give	a	slap	to	the	witch	green
	did not		a slap		by		green witch	
	no		slap		to the			
	did not give				to			
					the			
			slap			the witch		

- If we translate with beam search, what state do we need to keep in the beam?

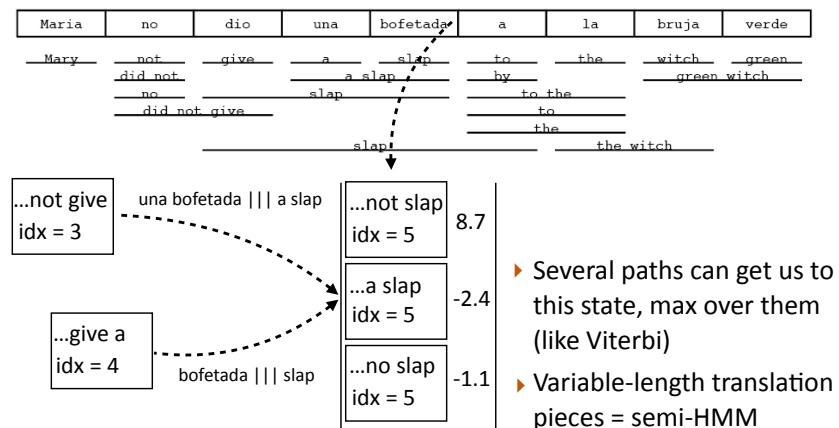
$$\arg \max_{\mathbf{e}} \left[\prod_{(\bar{e}, f)} P(\bar{f}|\bar{e}) \cdot \prod_{i=1}^{|\mathbf{e}|} P(e_i|e_{i-1}, e_{i-2}) \right]$$
- What have we translated so far?
- What words have we produced so far?
- When using a 3-gram LM, only need to remember the last 2 words!



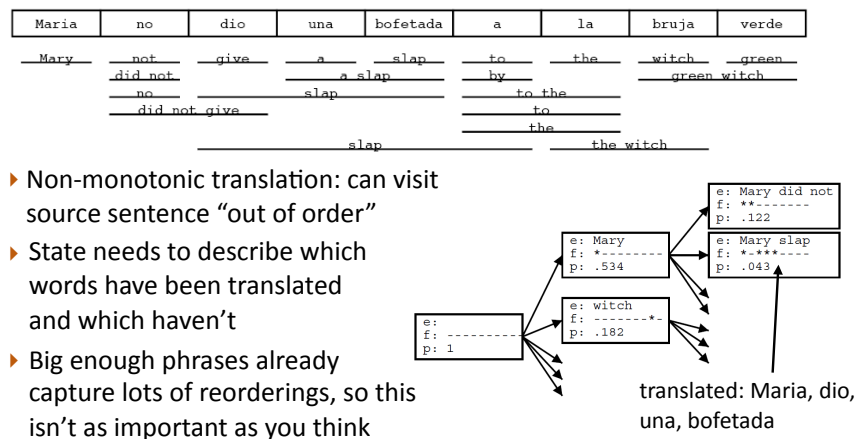
Monotonic Translation



Monotonic Translation



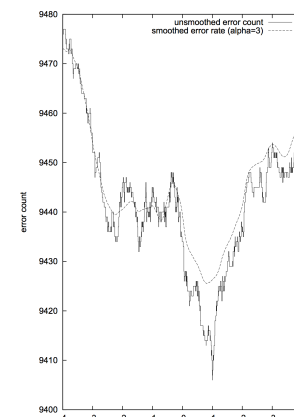
Non-Monotonic Translation



Training Decoders

score = $\alpha \log P(\mathbf{t}) + \beta \log P(\mathbf{s}|\mathbf{t})$
...and $P(\mathbf{s}|\mathbf{t})$ is in fact more complex

- Usually 5-20 feature weights to set, want to optimize for BLEU score which is not differentiable
- MERT (Och 2003): decode to get 1000-best translations for each sentence in a small training set (<1000 sentences), do line search on parameters to directly optimize for BLEU





Moses

- ▶ Toolkit for machine translation due to Philipp Koehn + Hieu Hoang
 - ▶ Pharaoh (Koehn, 2004) is the decoder from Koehn's thesis
- ▶ Moses implements word alignment, language models, and this decoder, plus *a ton* more stuff
 - ▶ Highly optimized and heavily engineered, could more or less build SOTA translation systems with this from 2007-2013



Moses

SOURCE	Cela constituerait une solution transitoire qui permettrait de conduire à terme à une charte à valeur contraignante.
HUMAN	That would be an interim solution which would make it possible to work towards a binding charter in the long term .
1x DATA	[this] [constituerait] [assistance] [transitoire] [who] [permettrait] [licences] [to] [terme] [to] [a] [charter] [to] [value] [contraignante] [.]
10x DATA	[it] [would] [a solution] [transitional] [which] [would] [of] [lead] [to] [term] [to a] [charter] [to] [value] [binding] [.]
100x DATA	[this] [would be] [a transitional solution] [which would] [lead to] [a charter] [legally binding] [.]
1000x DATA	[that would be] [a transitional solution] [which would] [eventually lead to] [a binding charter] [.]

slide credit:
Dan Klein



Syntactic MT

- ▶ Rather than use phrases, use a *synchronous context-free grammar*

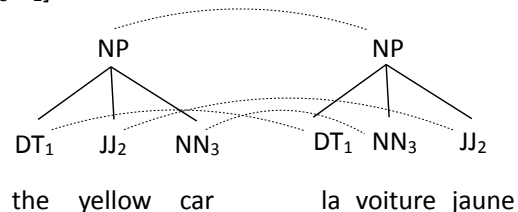
NP → [DT₁ JJ₂ NN₃; DT₁ NN₃ JJ₂]

DT → [the, la]

DT → [the, le]

NN → [car, voiture]

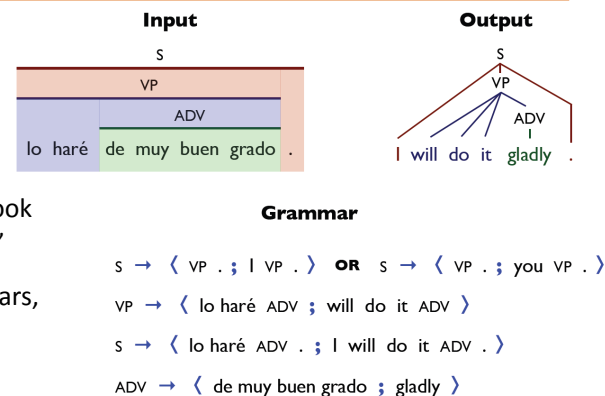
JJ → [yellow, jaune]



- ▶ Translation = parse the input with “half” of the grammar, read off the other half
- ▶ Assumes parallel syntax up to reordering



Syntactic MT



Slide credit: Dan Klein