Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

#### **Translation Options**



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

Dhillian Manha Halimaniko af Ediakona

4 4

Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

#### Hypothesis Expansion

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



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

- p: probability 0.534

16

Philipp Koehn, University of Edinburgh

10

Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

## Hypothesis Expansion

| Mary | not | give | a | slap | to | the | witch | green | slap | to | the | witch | green | slap | to | the | witch | green | slap | to | the | witch | green | slap | to | the | witch | green | slap | to | the | witch | green | slap | to | the | witch | green | slap | to | the | witch | green | slap | to | the | witch | green | slap | the | witch | green | the | witch | the | witch | green | the | witch |

• Further hypothesis expansion

Dhillian Manha Halisasaiks of Ediahssash

\_

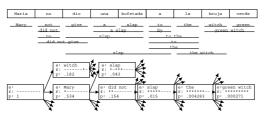
S

E

F

Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

## Hypothesis Expansion



- Adding more hypothesis
- $\Rightarrow$  Explosion of search space

Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

#### Hypothesis Expansion





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

Philipp Koehn, University of Edinburgh

15

F

Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

# Hypothesis Expansion



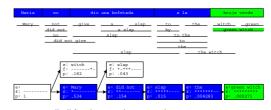
• Add another hypothesis

Philipp Koehn University of Edinburgh

18

Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

## Hypothesis Expansion



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

Dhilles Mache Halisasaks of Ediahssa

-00

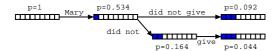
Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

## **Explosion of Search Space**

- Number of hypotheses is exponential with respect to sentence length
- $\Rightarrow$  Decoding is NP-complete [Knight, 1999]
- $\Rightarrow$  Need to reduce search space
  - risk free: hypothesis recombination
  - risky: histogram/threshold pruning

Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

# Hypothesis Recombination



• Different paths to the same partial translation

Philipp Koehn, University of Edinburgh

23

F

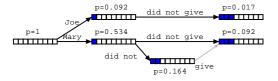
S

\_

S

Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

# 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)

Dhiling Macha Hairmaik, of Ediahrosah

٥r

Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

# **Pruning**

- Hypothesis recombination is not sufficient
- ⇒ Heuristically discard weak hypotheses
- Organize Hypothesis in stacks, e.g. by
  - same foreign words covered
  - same number of foreign words covered (Pharaoh does this)
  - 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  $\alpha$  times the cost of best hypothesis in stack (e.g.,  $\alpha$  = 0.001)

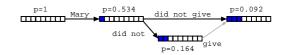
Philipp Koehn, University of Edinburgh

27

F

Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

# Hypothesis Recombination



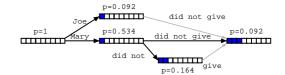
- Different paths to the same partial translation
- $\Rightarrow$  Combine paths
  - drop weaker hypothesis
  - keep pointer from worse path

Philipp Koehn University of Edinburgh

24

#### Statistical Machine Translation — Lecture 2: Theory and Praxis of Decoding

#### 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

Dhillian Manha Thiringaist, of Ediah, sah