Statistical Machine Translation

Lecture 4

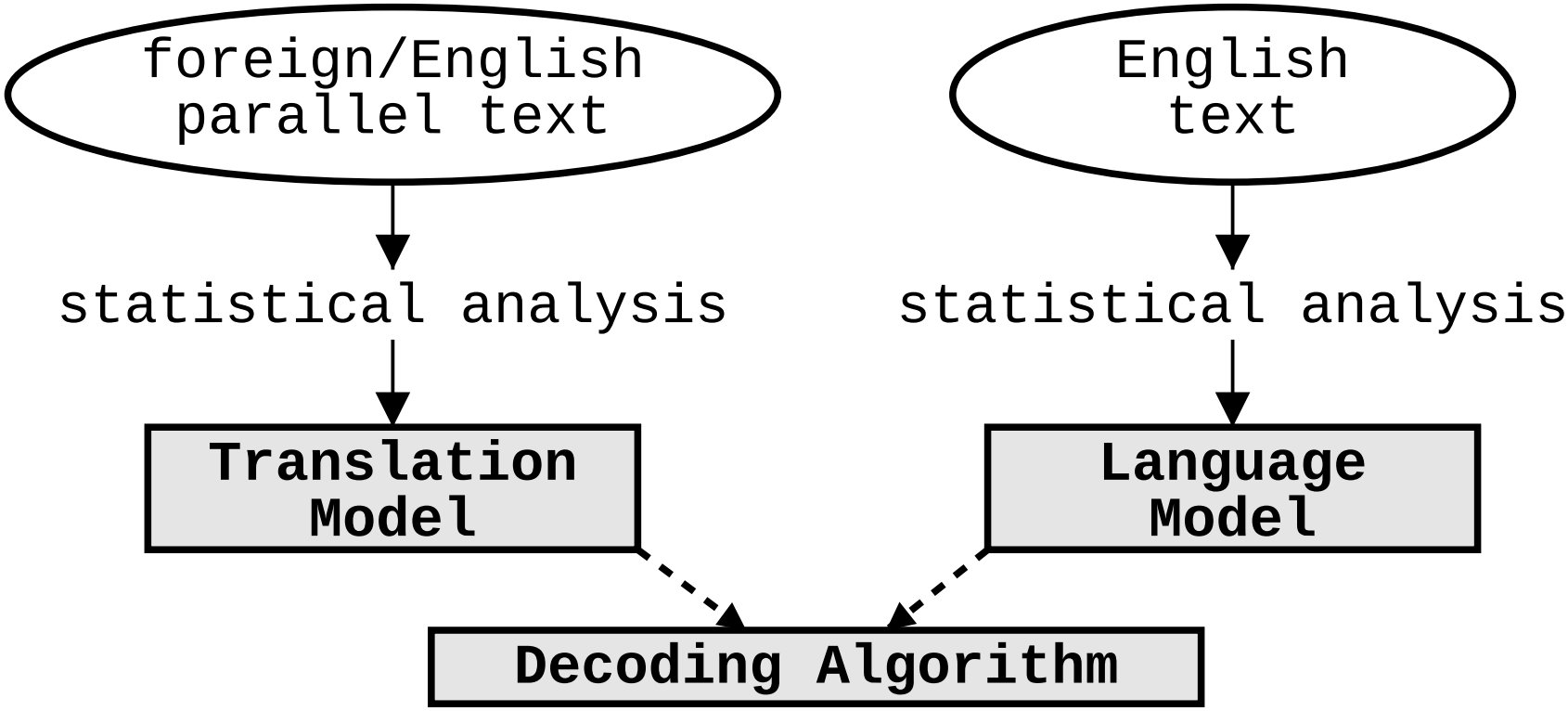
Decoding with Phrase-Based Models

Stephen Clark

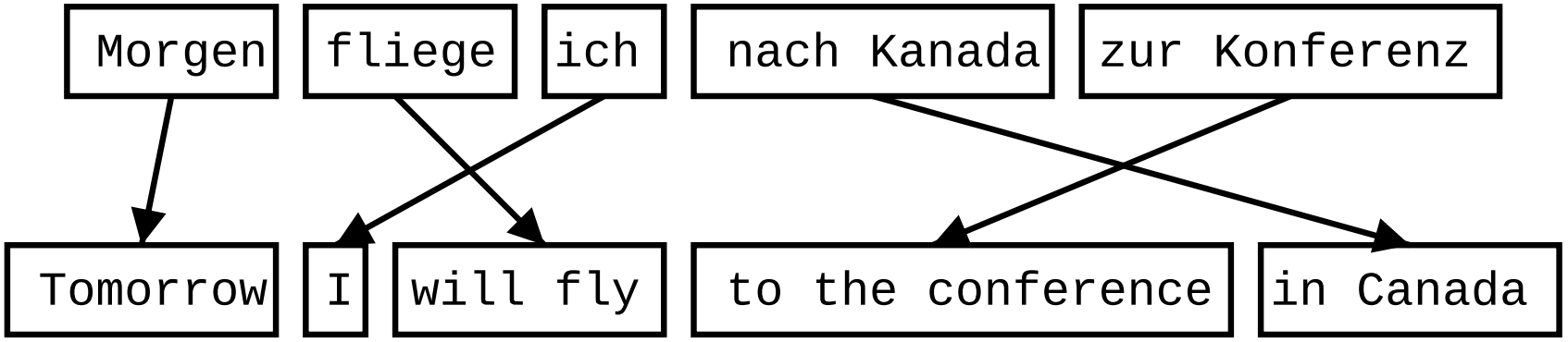
(based on slides by Phillip Koehn)

Statistical Machine Translation

- Components: Translation model, language model, decoder



Phrase-Based Translation



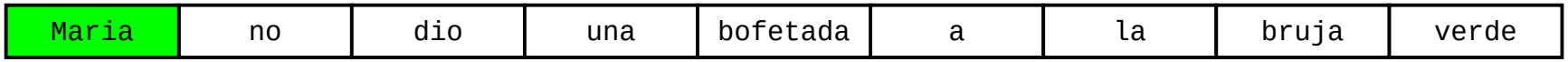
- Foreign input is segmented in phrases
  - any sequence of words, not necessarily linguistically motivated
- Each phrase is translated into English
- Phrases are reordered

Phrase Translation Table

- Phrase Translations for “den Vorschlag”:

English	ϕ**(e|f)**	**English**	ϕ**(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



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

Decoding Process

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

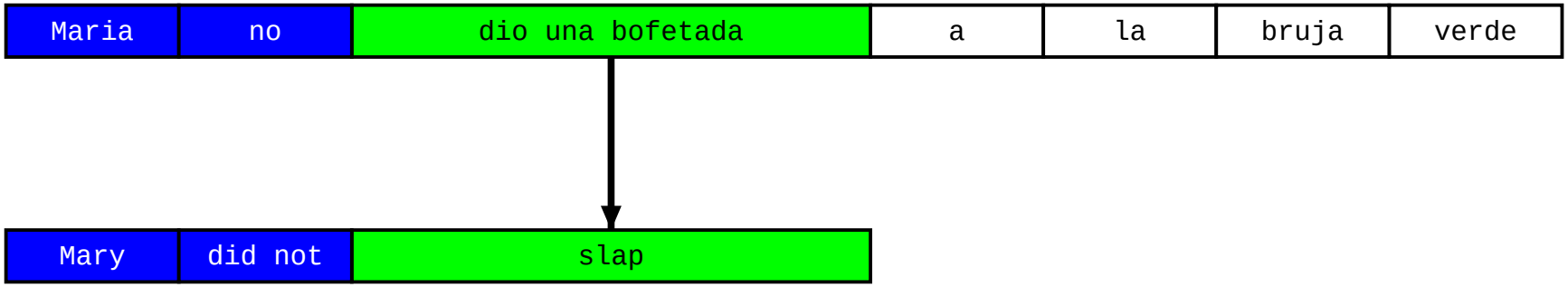
Decoding Process

- 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

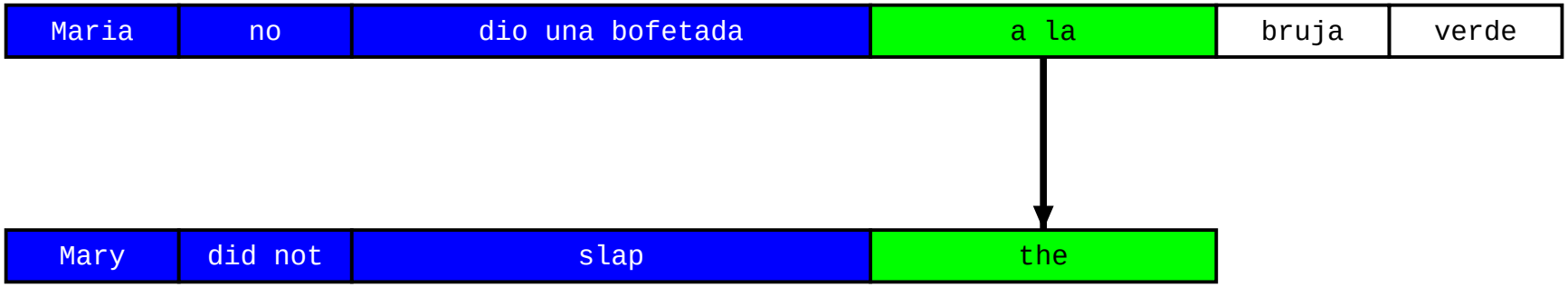
- One to many translation

Decoding Process



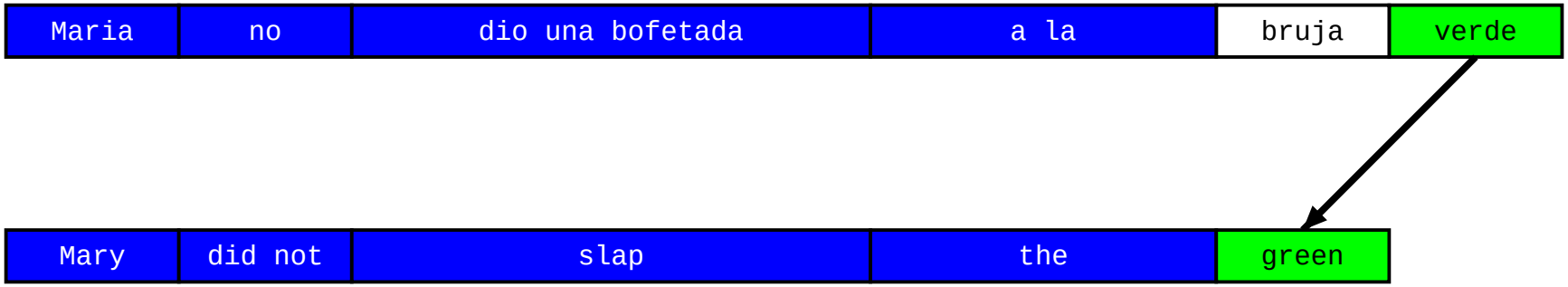
- Many to one translation

Decoding Process



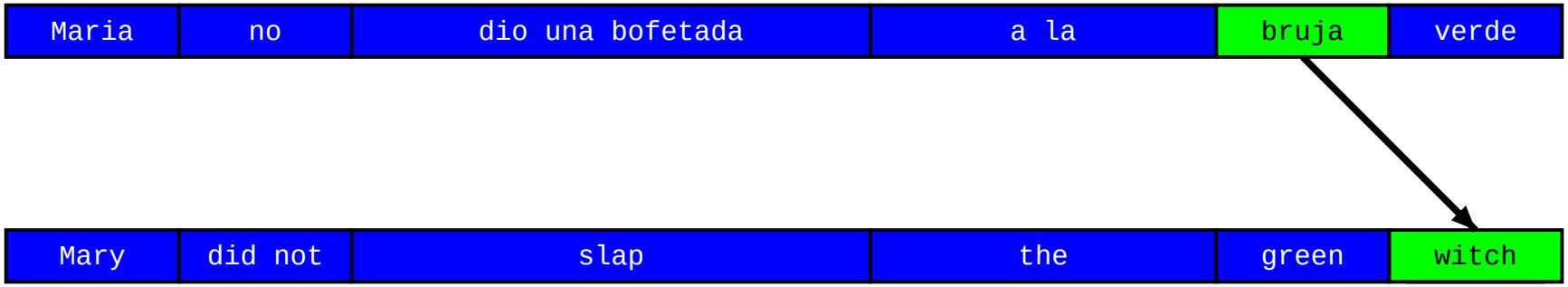
- Many to one translation

Decoding Process



- Reordering

Decoding Process

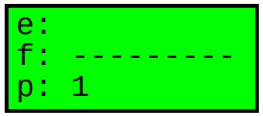


- Translation finished

Translation Options

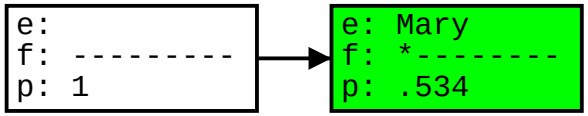
- Look up possible phrase translations
  - many different ways to segment words into phrases
  - many different ways to translate each phrase

Hypothesis Expansion



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

Hypothesis Expansion

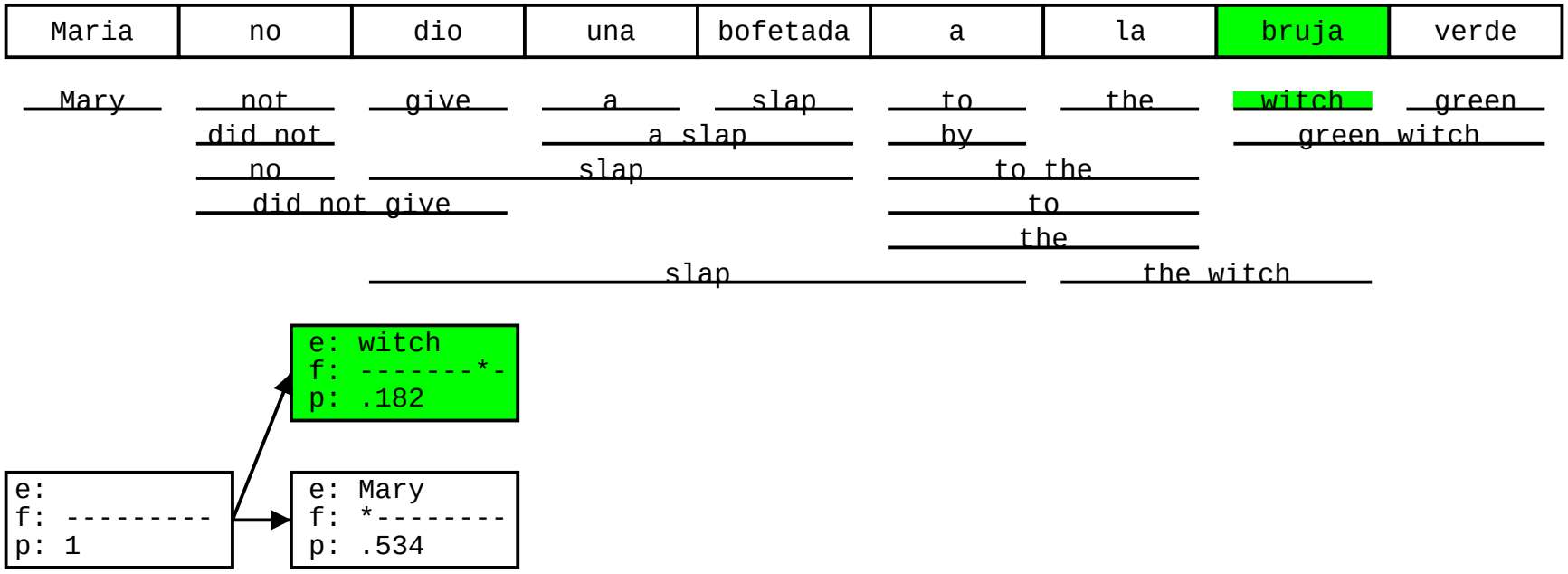


- Pick translation option
- Create hypothesis
  - e: add English phrase `Mary`
  - f: first foreign word covered
  - p: probability 0.534

A Quick Word on Probabilities

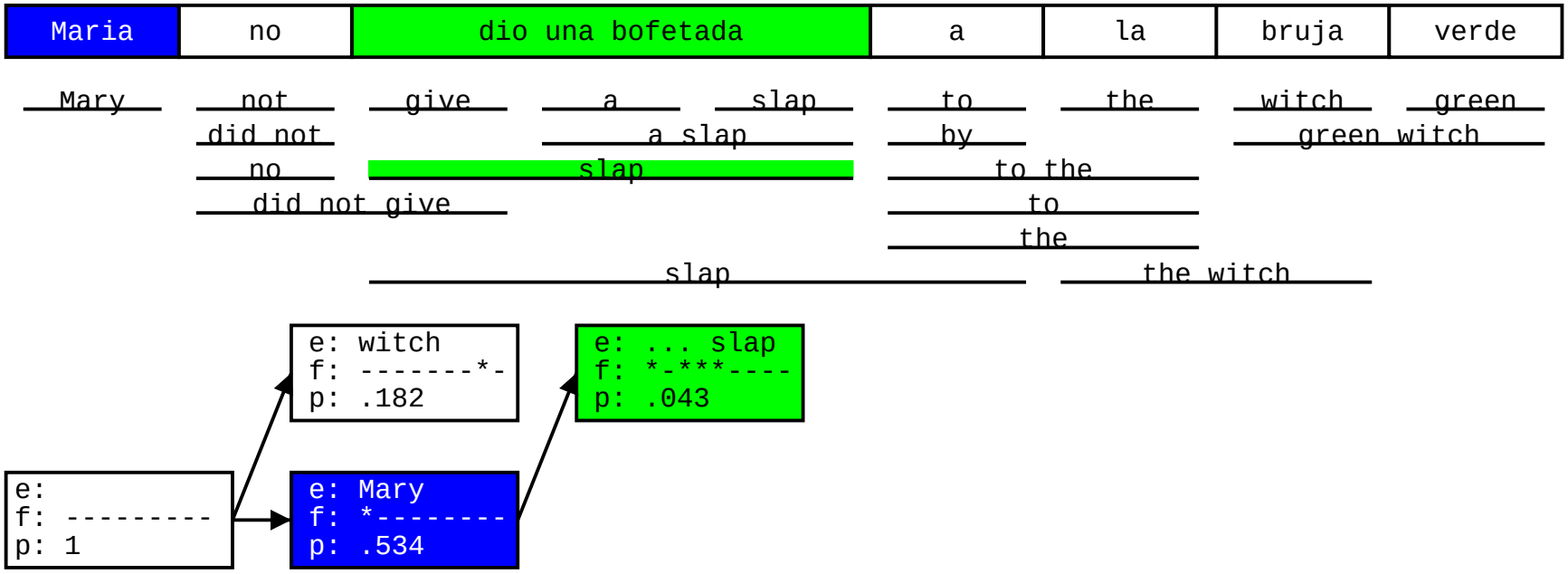
- Not going into detail here, but...
- Translation Model
  - phrase translation probability p(Mary|Maria)
  - reordering costs
  - phrase/word count costs
  - ...
- Language Model
  - uses trigrams:
  - p(Mary did not) = p(Mary| <s>) * p(did|Mary,<s>) * p(not|Mary did)

Hypothesis Expansion



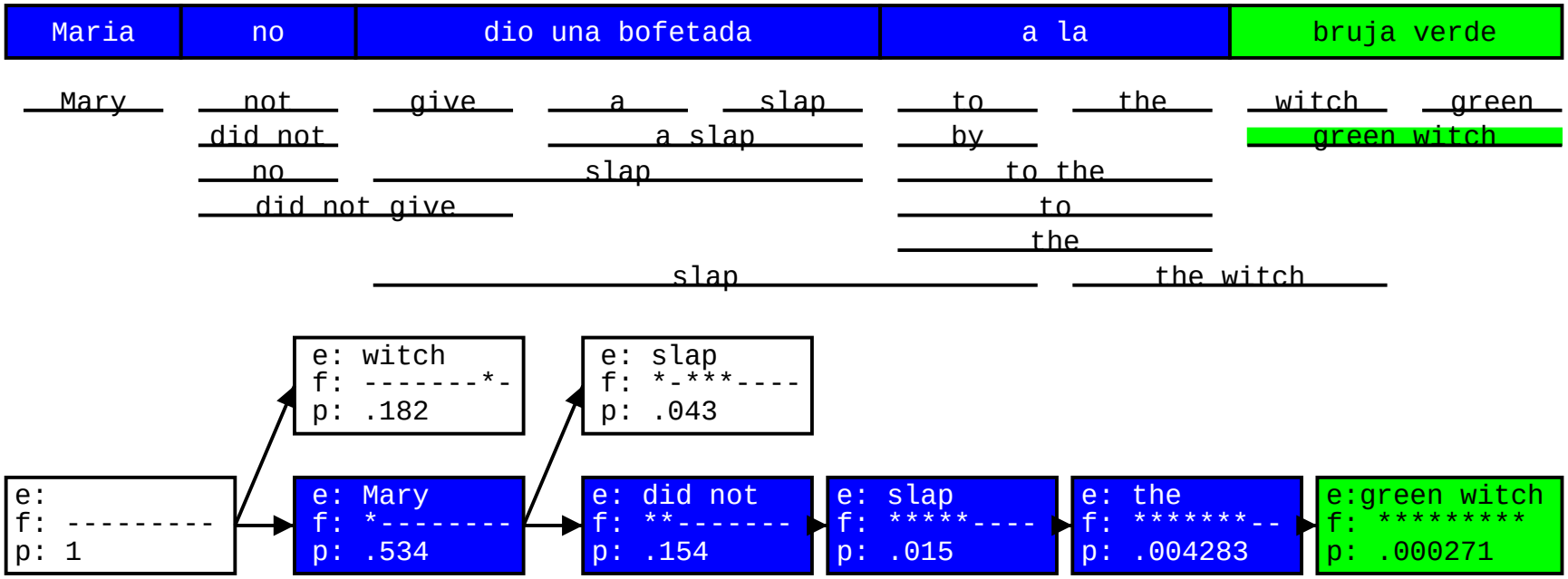
- Add another hypothesis

Hypothesis Expansion



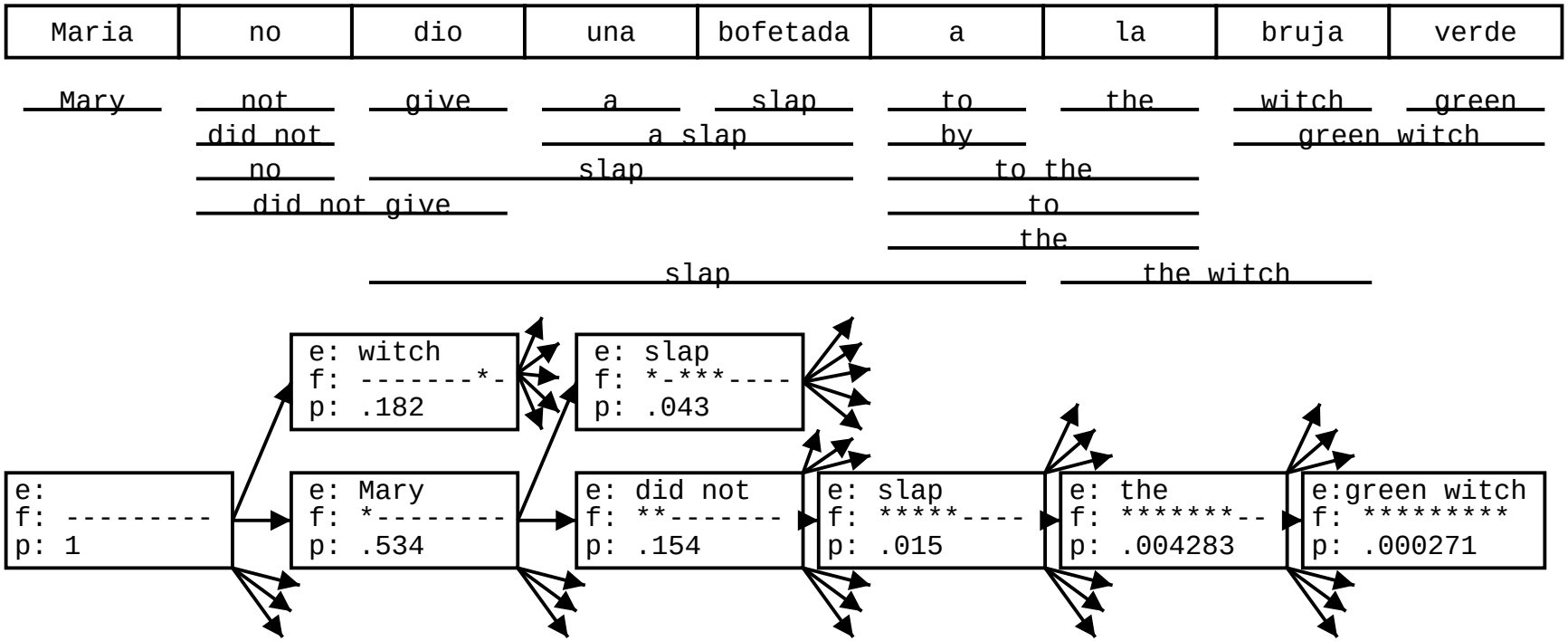
- Further hypothesis expansion

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

$\Rightarrow$ Explosion of search space

Explosion of Search Space

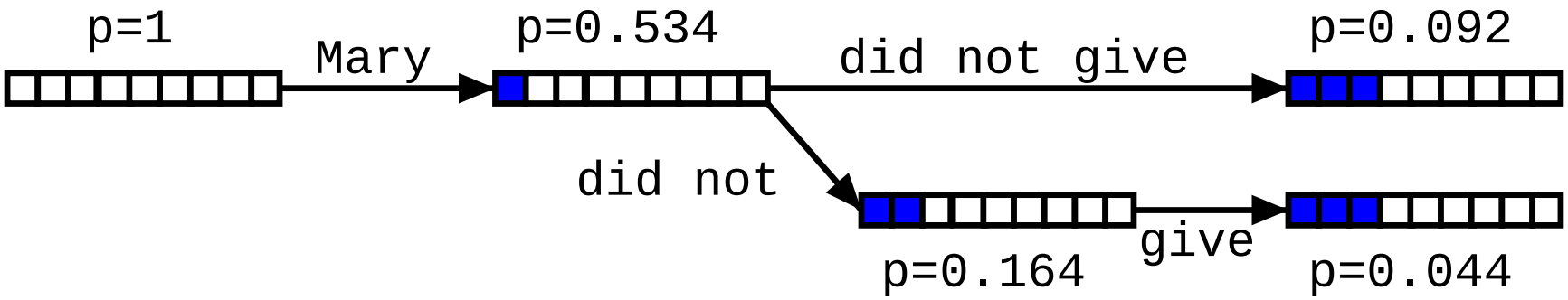
- Number of hypotheses is exponential with respect to sentence length

$\Rightarrow$ Decoding is NP-complete [Knight, 1999]

$\Rightarrow$ Need to reduce search space

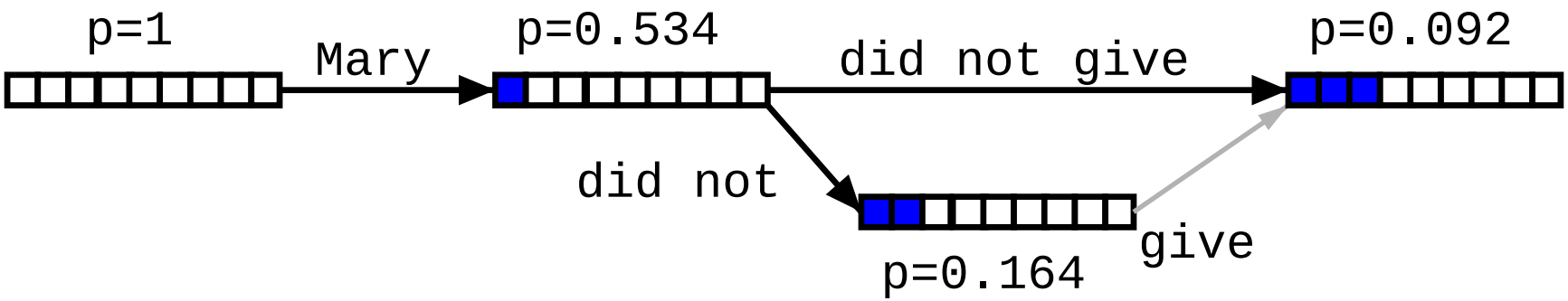
  - risk free: hypothesis recombination
  - risky: histogram/threshold pruning

Hypothesis Recombination



- Different paths to the same partial translation

Hypothesis Recombination

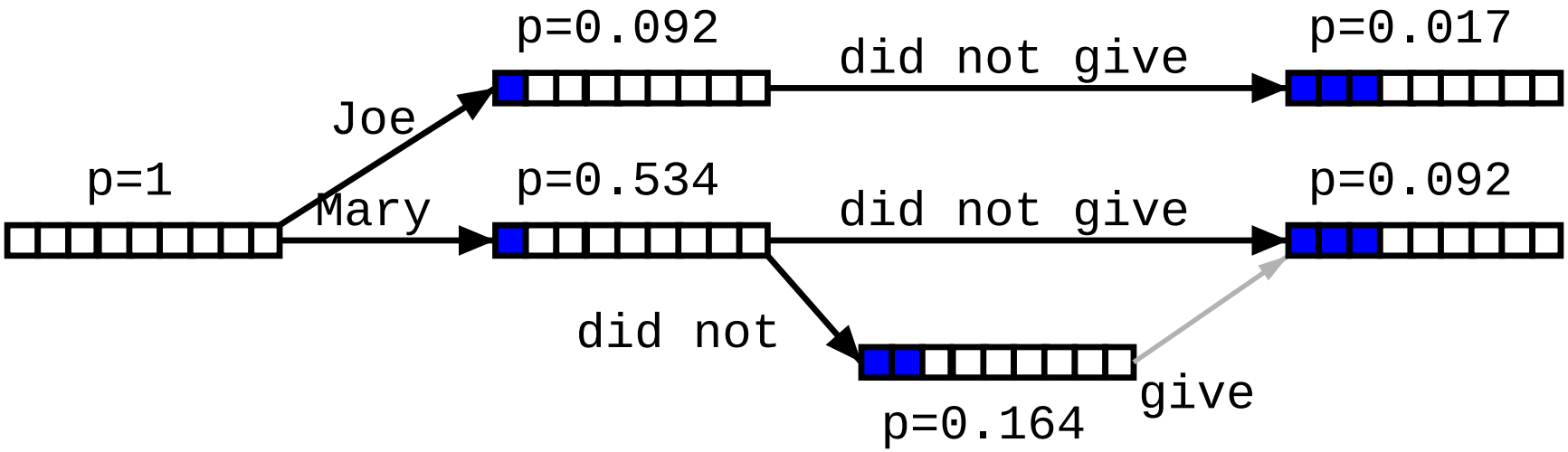


- Different paths to the same partial translation

$\Rightarrow$ Combine paths

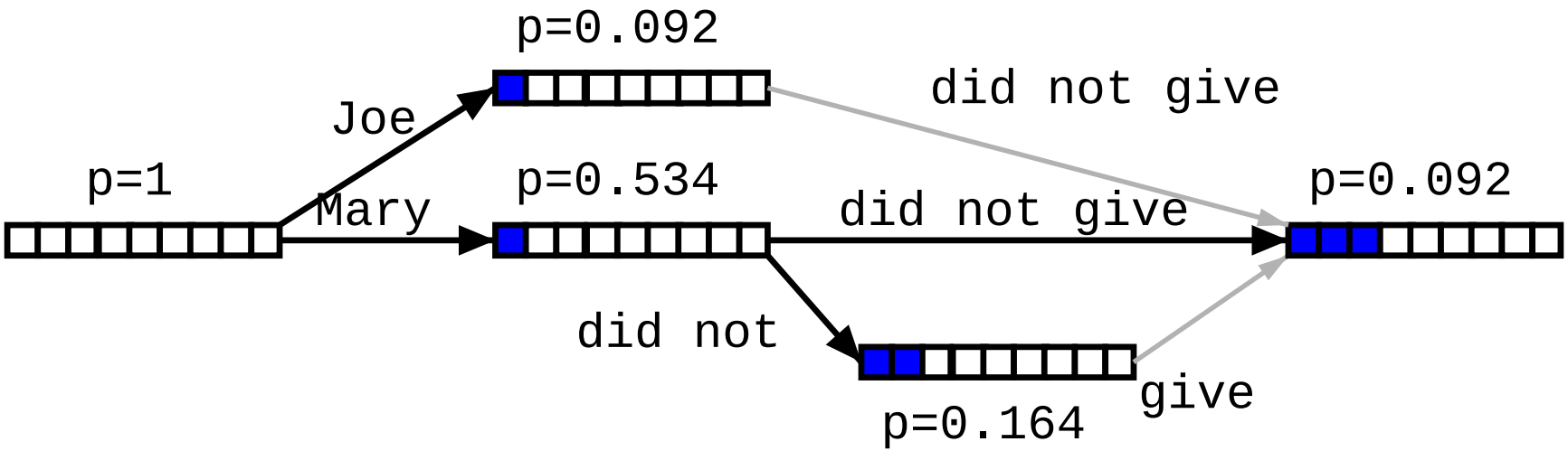
  - drop weaker hypothesis
  - keep pointer from worse path

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 (affects 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)

$\Rightarrow$ Combine paths

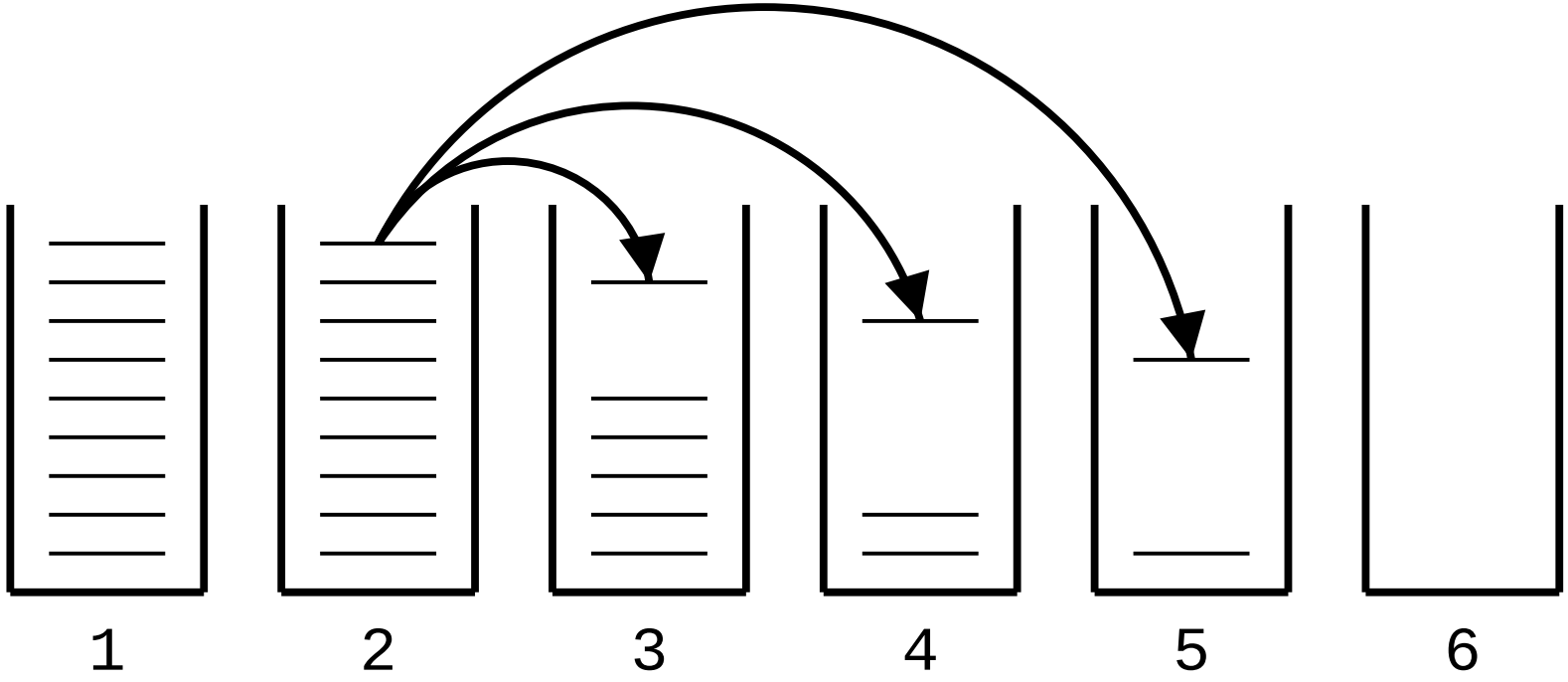
Pruning

- Hypothesis recombination is not sufficient

$\Rightarrow$ Heuristically discard weak hypotheses

- 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., α = 0.001)

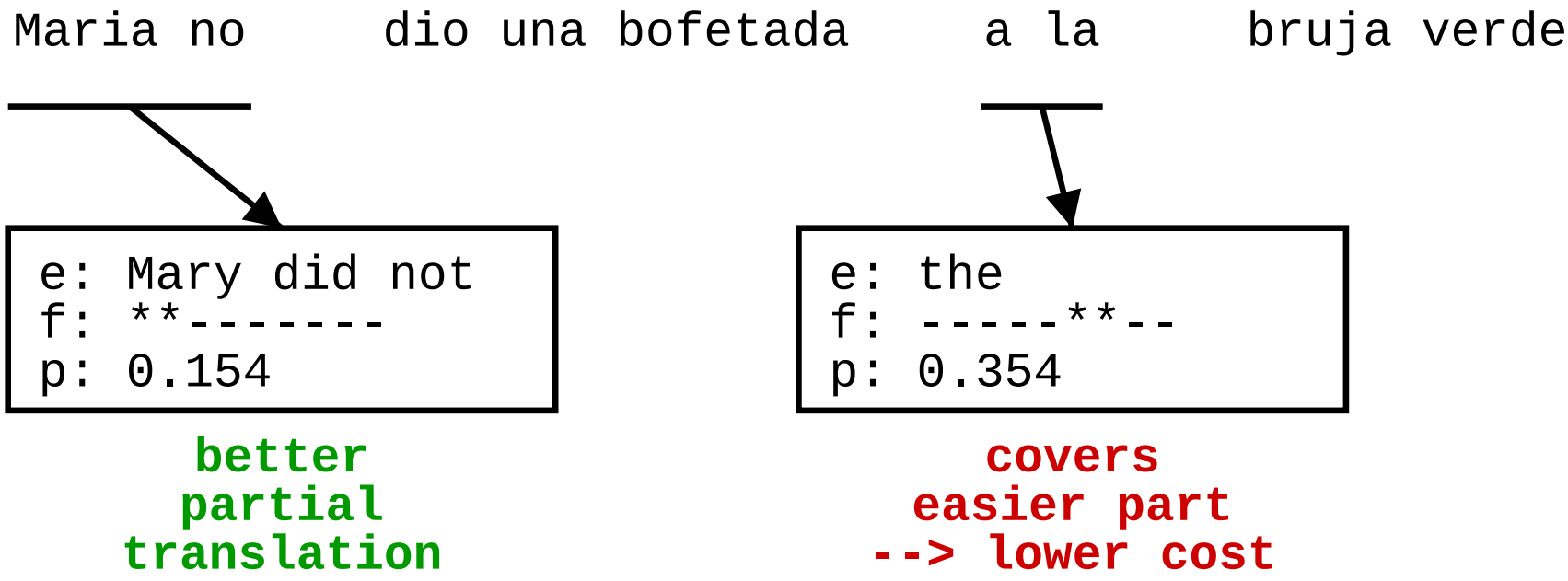
Hypothesis Stacks



- 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



- Hypothesis that covers *easy* part of sentence is preferred

$\Rightarrow$ Need to consider future cost of uncovered parts