Machine translation: Decoding

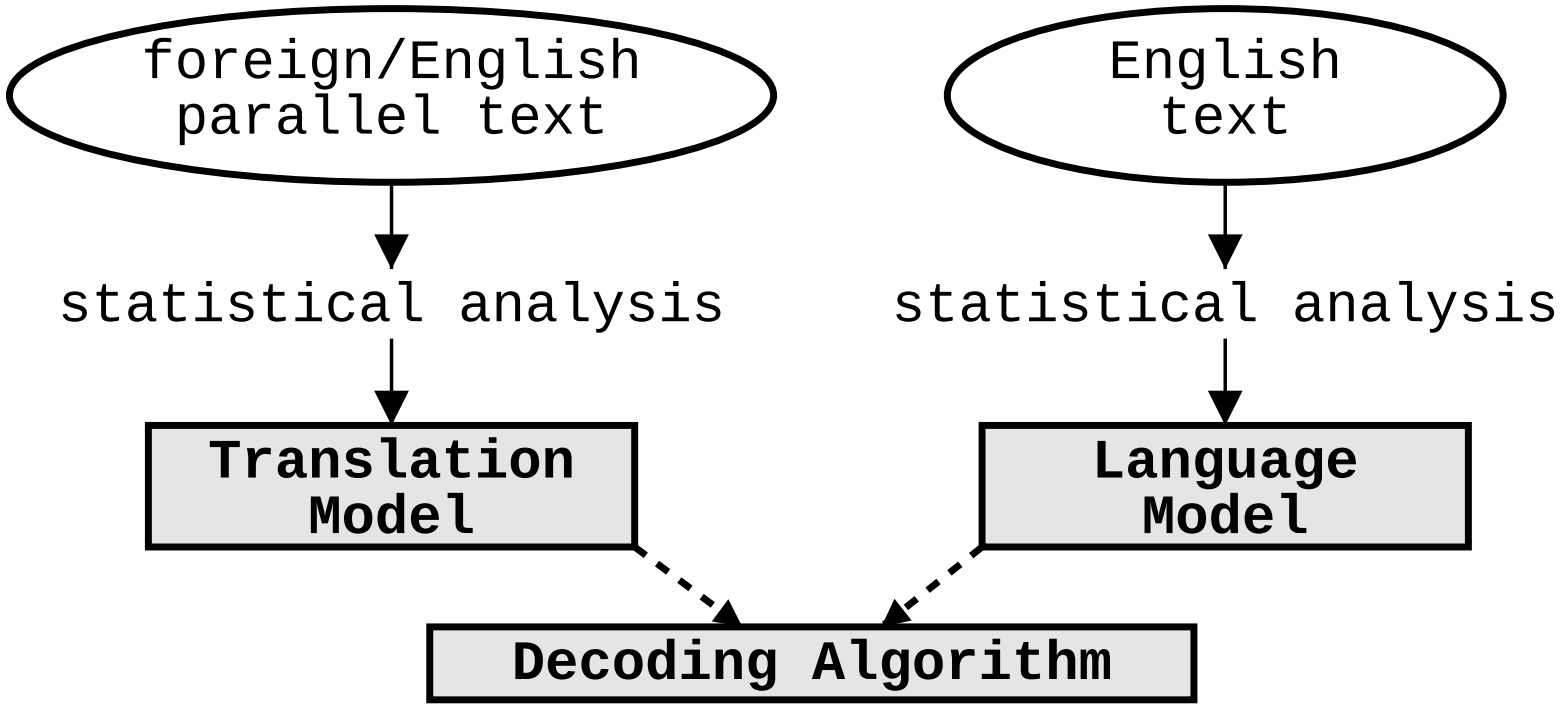
Chris Callison-Burch
Slides borrowed from Philipp Koehn

December 4, 2007

JOHNS HOPKINS UNIVERSITY

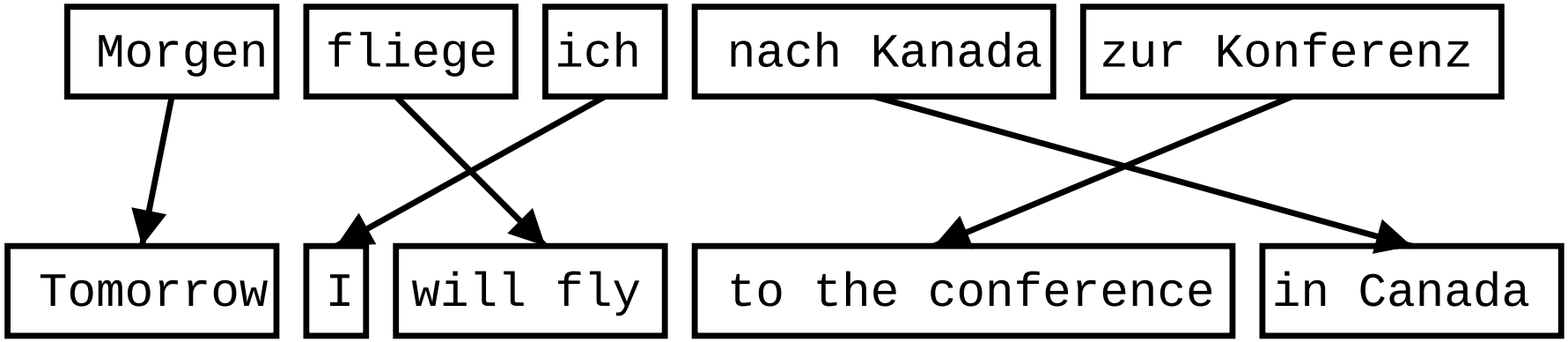
Statistical Machine Translation

- Components: Translation model, language model, decoder



JOHNS HOPKINS UNIVERSITY

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

JOHNS HOPKINS UNIVERSITY

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	...	...

JOHNS HOPKINS UNIVERSITY

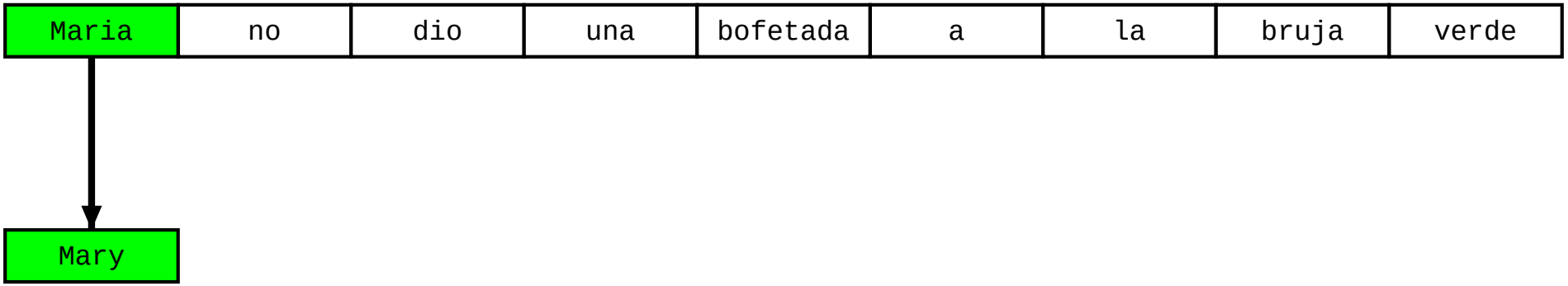
Decoding Process

Maria	no	dio	una	bofetada	a	la	bruja	verde

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

JOHNS HOPKINS UNIVERSITY

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

JOHNS HOPKINS UNIVERSITY

Decoding Process

Maria	no	dio	una	bofetada	a	la	bruja	verde

Mary

- 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

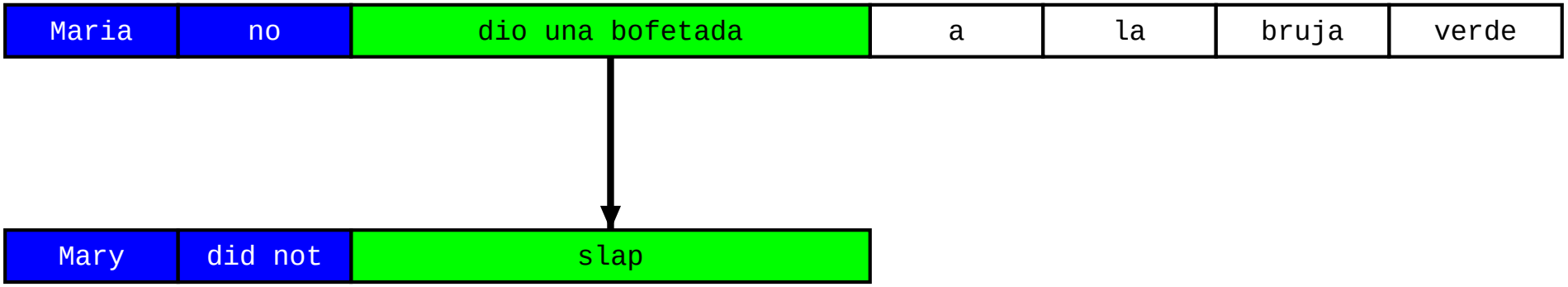
JOHNS HOPKINS UNIVERSITY

Decoding Process

- *One to many* translation

JOHNS HOPKINS UNIVERSITY

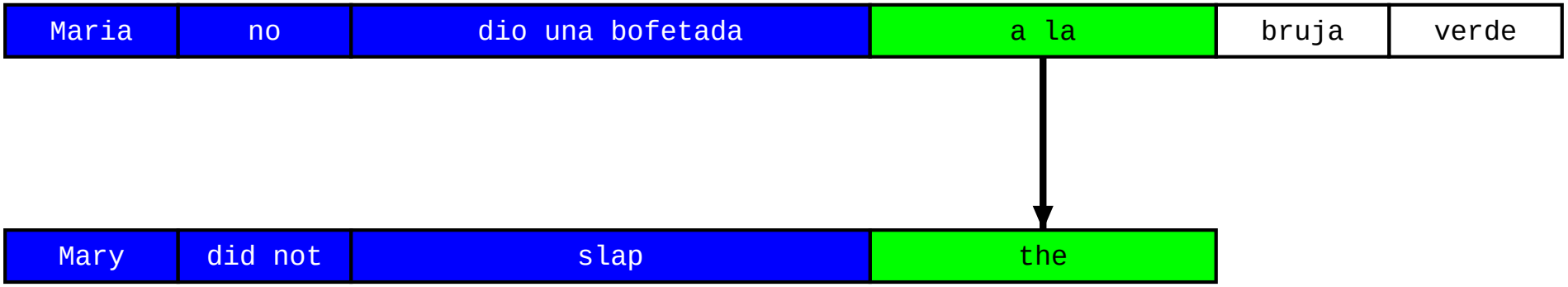
Decoding Process



- Many to one translation

JOHNS HOPKINS UNIVERSITY

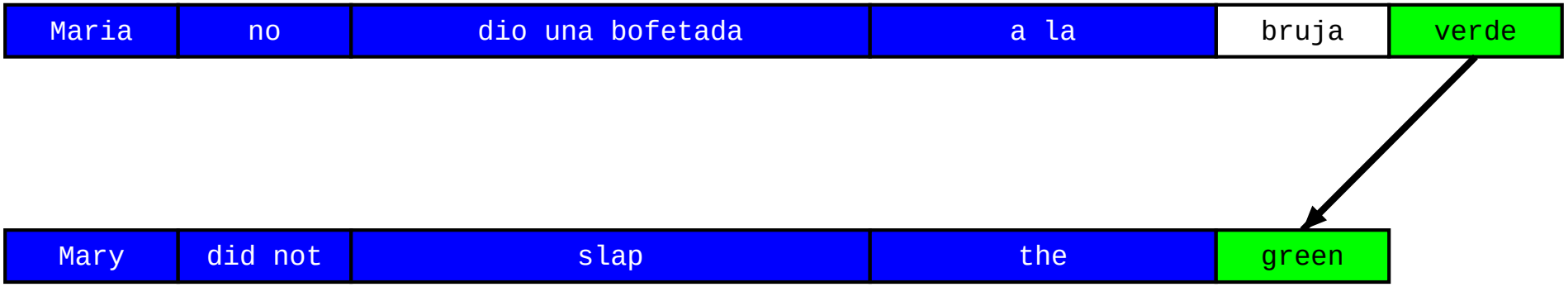
Decoding Process



- *Many to one* translation

JOHNS HOPKINS UNIVERSITY

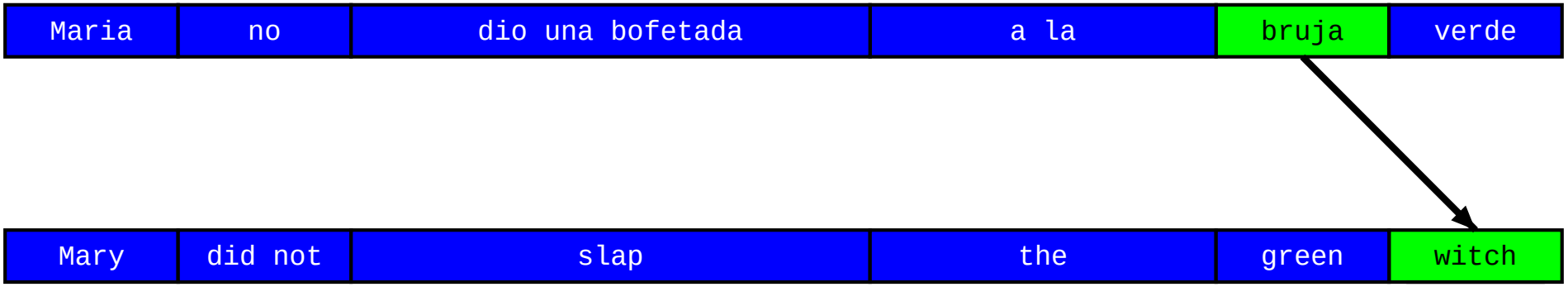
Decoding Process



- *Reordering*

JOHNS HOPKINS UNIVERSITY

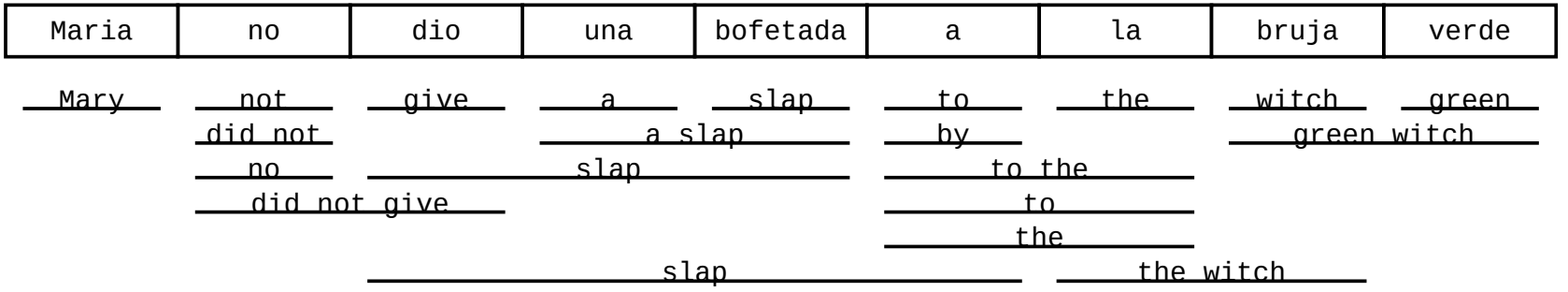
Decoding Process



- Translation *finished*

JOHNS HOPKINS UNIVERSITY

Translation Options



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

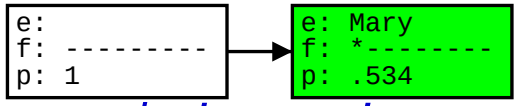
JOHNS HOPKINS UNIVERSITY

Hypothesis Expansion

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

JOHNS HOPKINS UNIVERSITY

Hypothesis Expansion



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

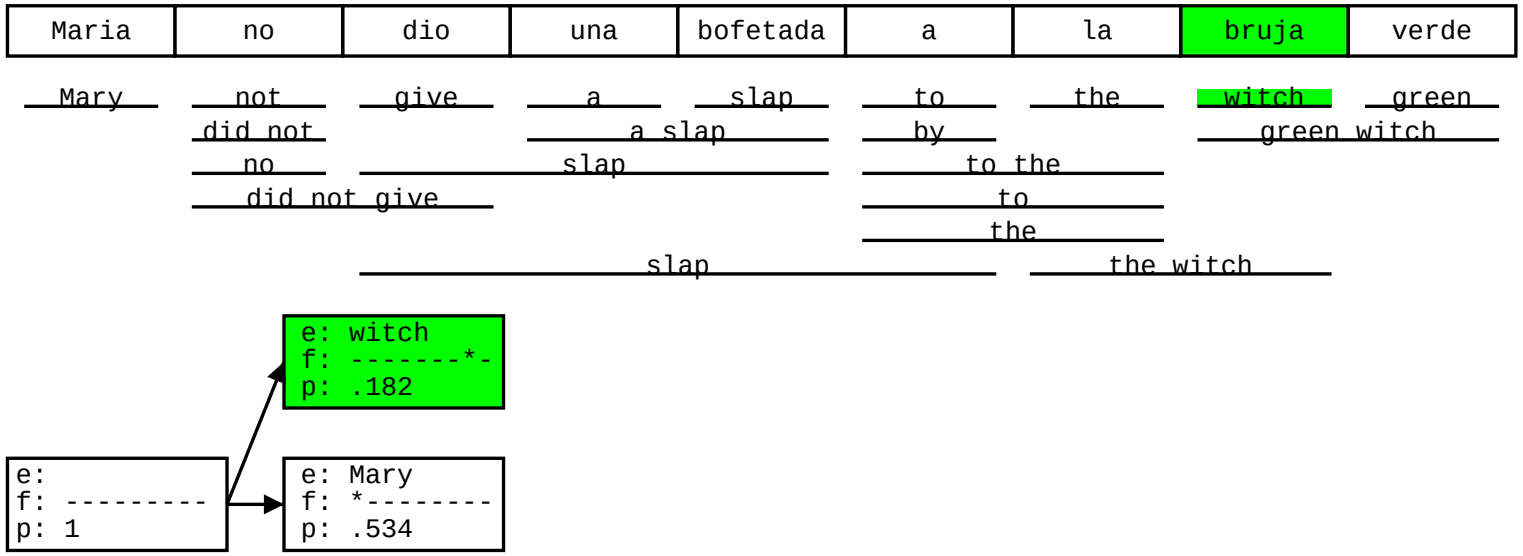
JOHNS HOPKINS UNIVERSITY

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|START) $\times p$(did|Mary,START) $\times$ p(not|Mary did)

JOHNS HOPKINS UNIVERSITY

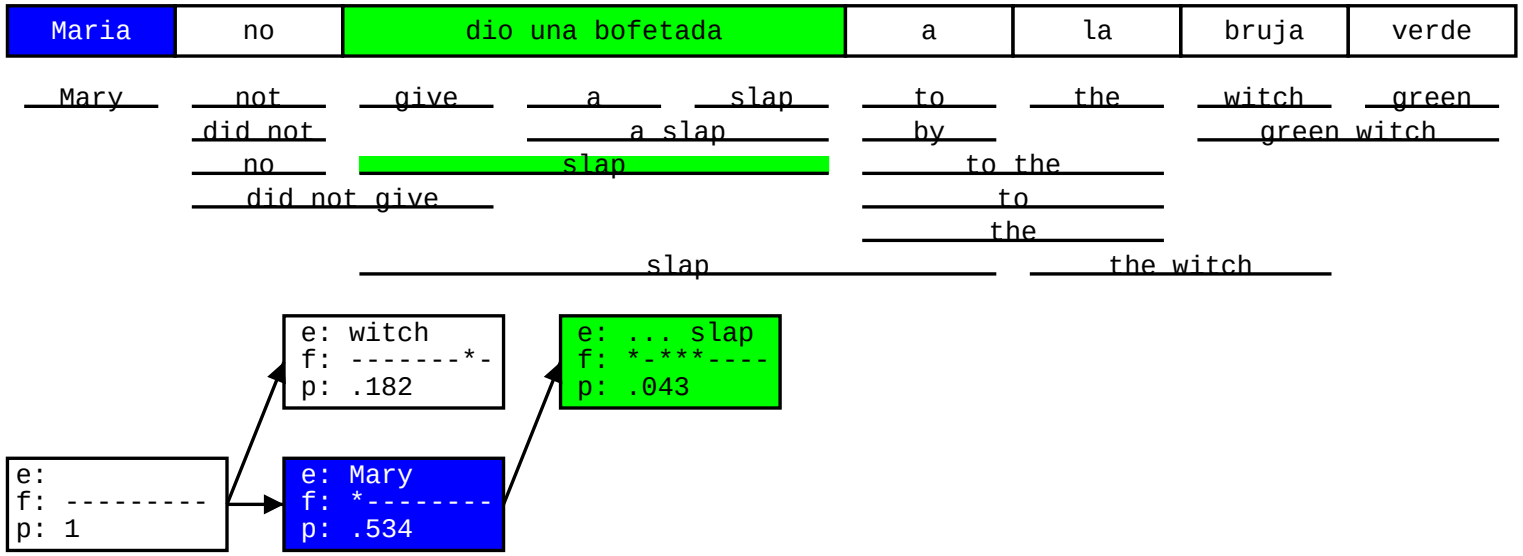
Hypothesis Expansion



- Add another *hypothesis*

JOHNS HOPKINS UNIVERSITY

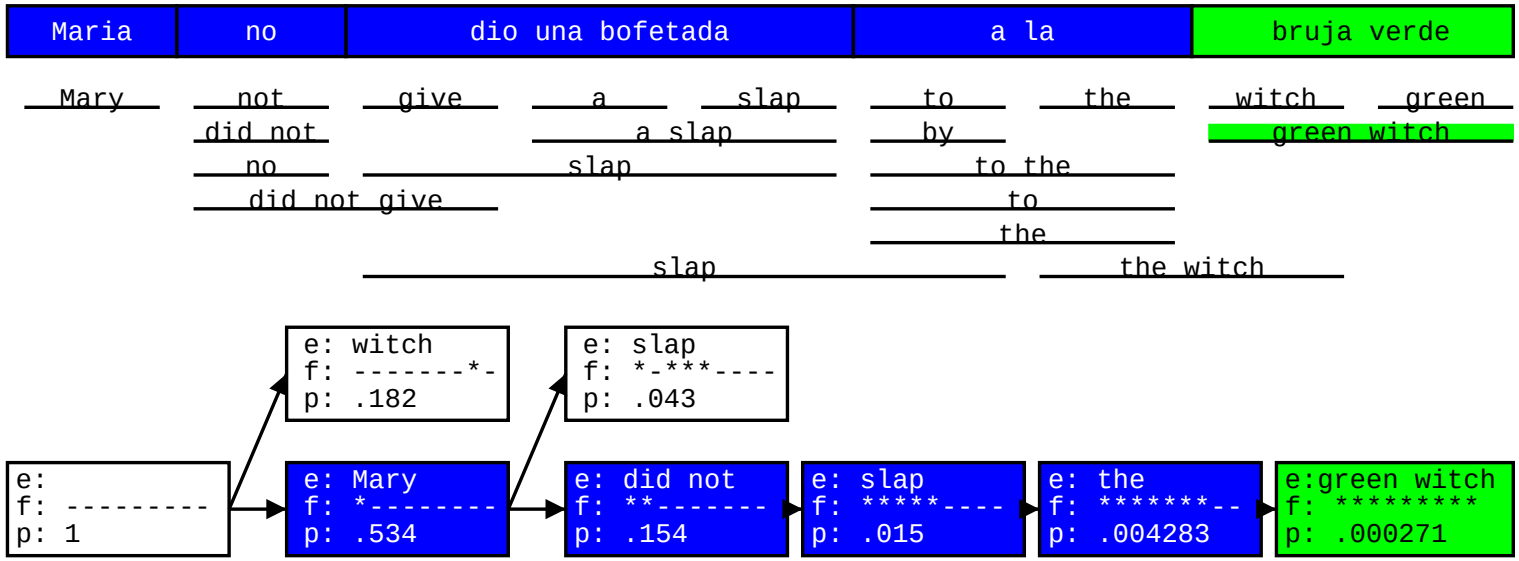
Hypothesis Expansion



- Further *hypothesis expansion*

JOHNS HOPKINS UNIVERSITY

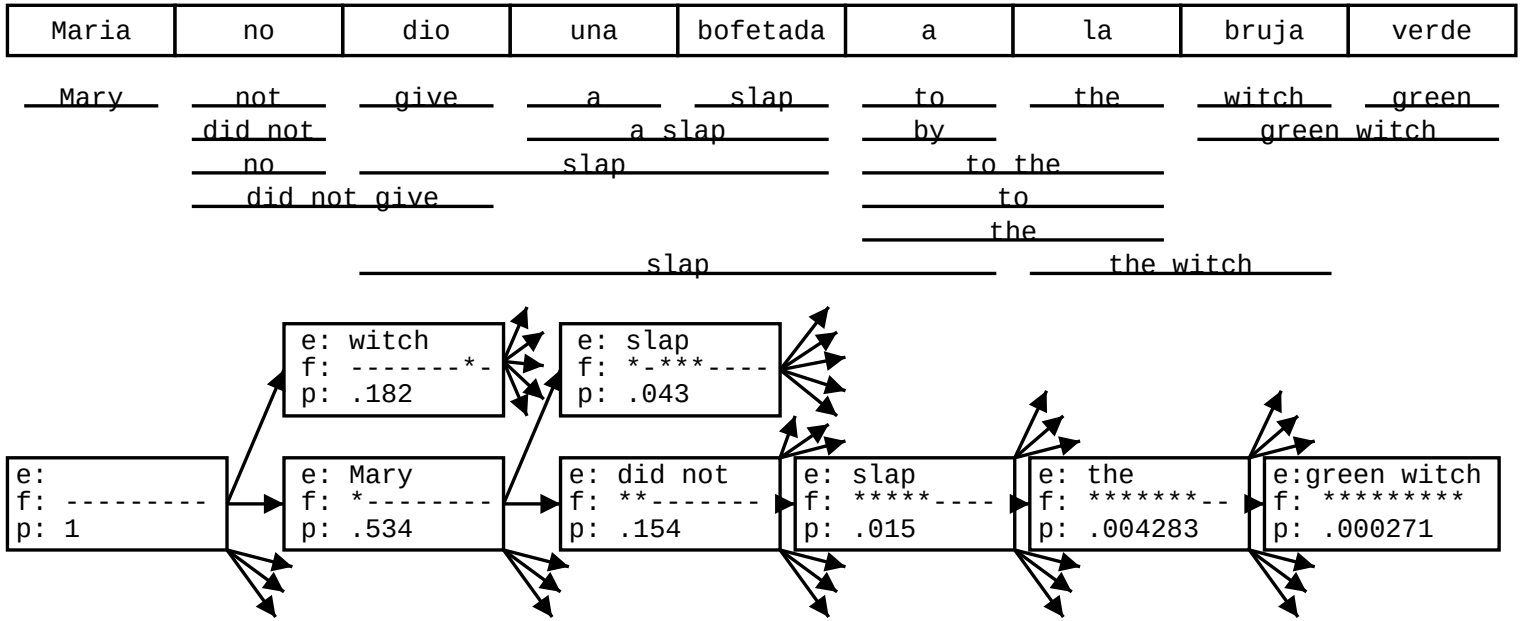
Hypothesis Expansion



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

JOHNS HOPKINS UNIVERSITY

Hypothesis Expansion



- Adding more hypothesis

$\Rightarrow$ *Explosion* of search space

JOHNS HOPKINS UNIVERSITY

Explosion of Search Space

- Number of hypotheses is *exponential* with respect to sentence length

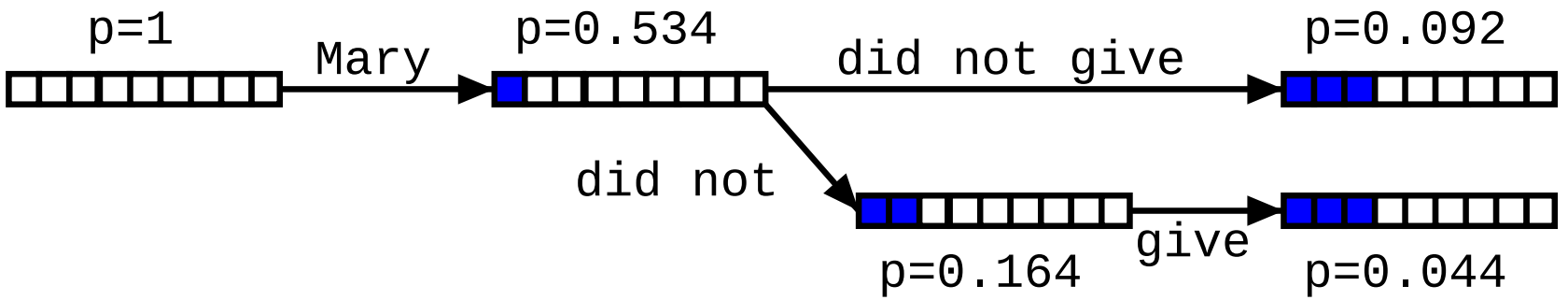
$\Rightarrow$ Decoding is NP-complete [Knight, 1999]

$\Rightarrow$ Need to *reduce search space*

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

JOHNS HOPKINS UNIVERSITY

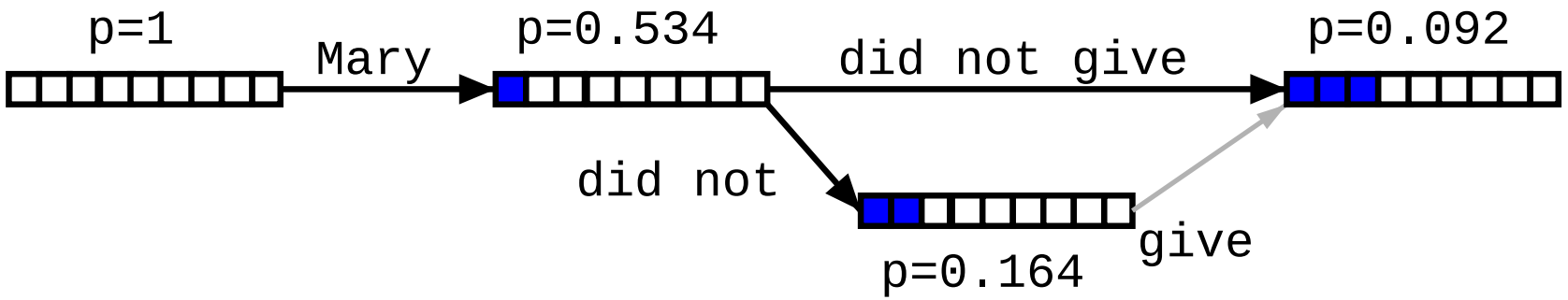
Hypothesis Recombination



- Different paths to the *same* partial translation

JOHNS HOPKINS UNIVERSITY

Hypothesis Recombination



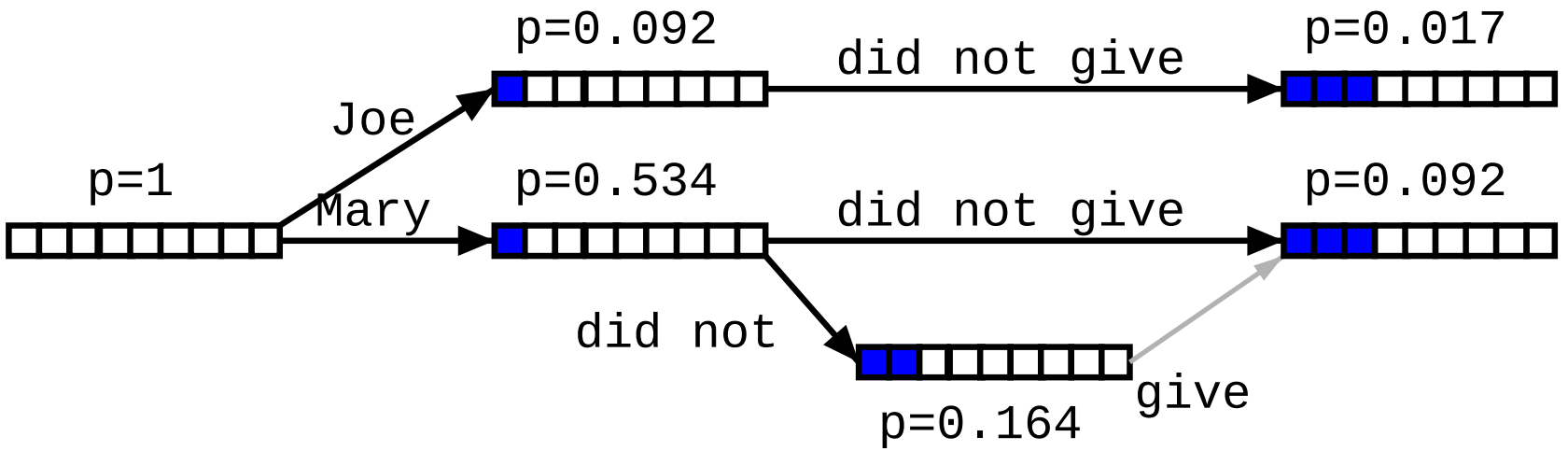
- Different paths to the same partial translation

⇒ *Combine paths*

  - *drop weaker* path
  - keep pointer from weaker path (for lattice generation)

JOHNS HOPKINS UNIVERSITY

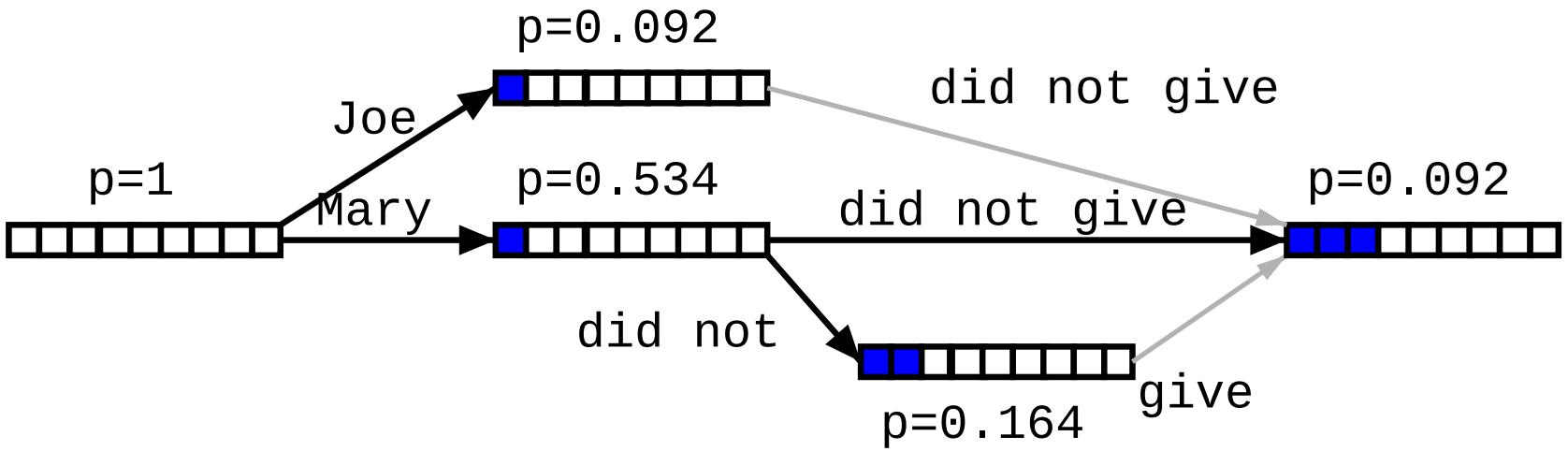
Hypothesis Recombination



- 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)

JOHNS HOPKINS UNIVERSITY

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*

JOHNS HOPKINS UNIVERSITY

Pruning

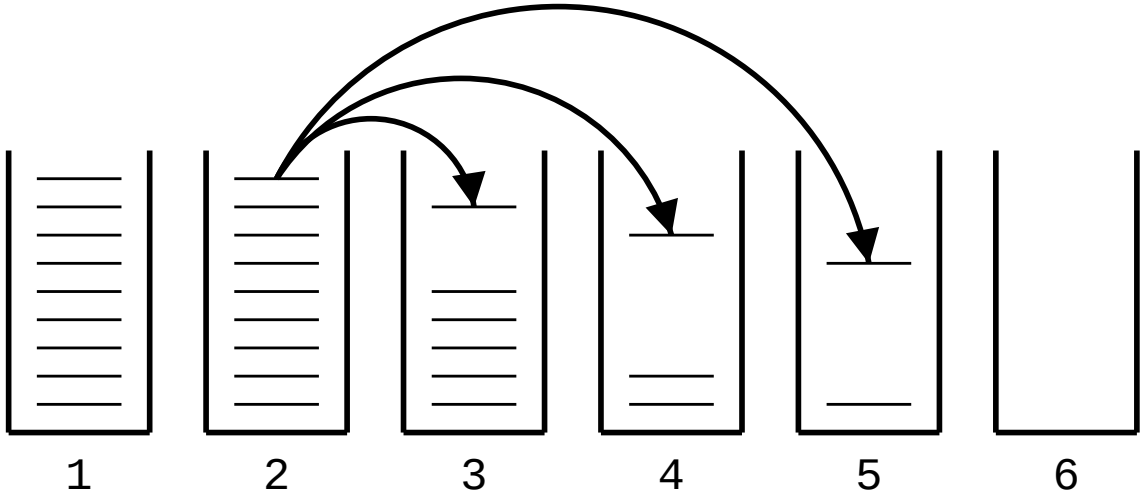
- Hypothesis recombination is *not sufficient*

⇒ 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., $\alpha = 0.001$)

JOHNS HOPKINS UNIVERSITY

Hypothesis Stacks

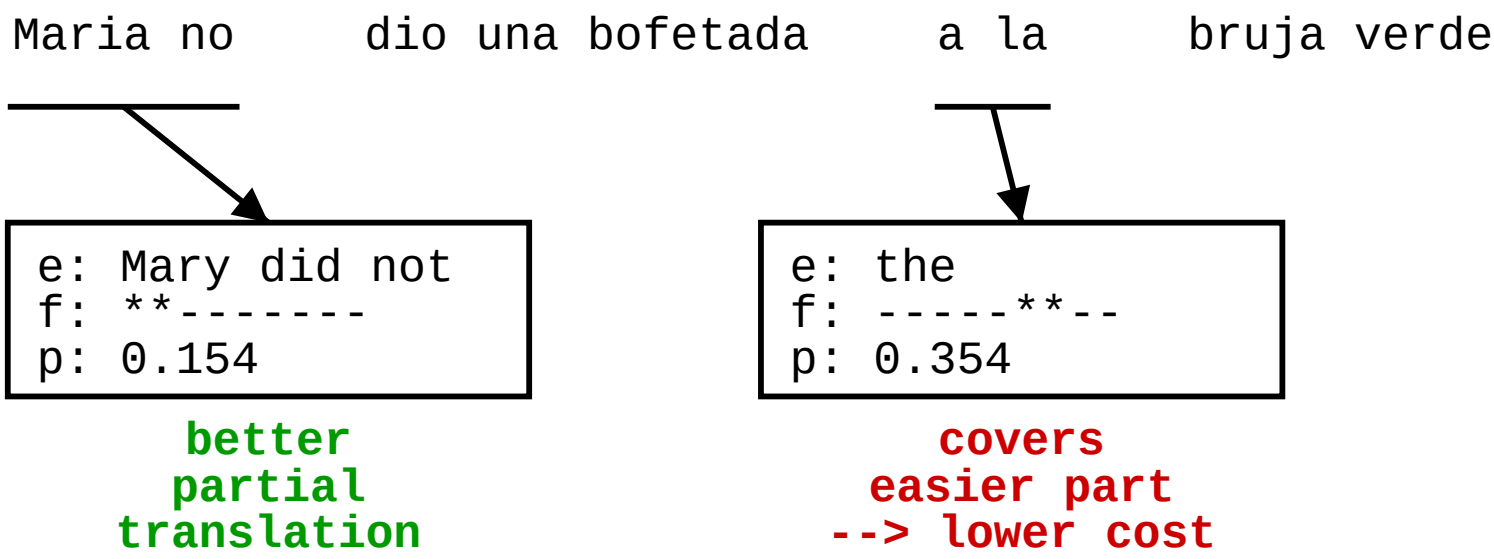


- Organization of hypothesis into stacks
  - here: based on *number of foreign words* translated
  - during translation all hypotheses from one stack are expanded
  - expanded Hypotheses are placed into stacks

JOHNS HOPKINS UNIVERSITY

Comparing Hypotheses

- Comparing hypotheses with *same number of foreign words* covered

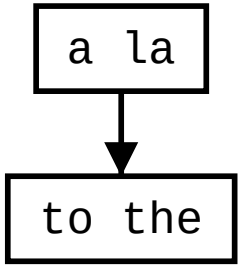


- Hypothesis that covers *easy part* of sentence is preferred

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

JOHNS HOPKINS UNIVERSITY

Future Cost Estimation

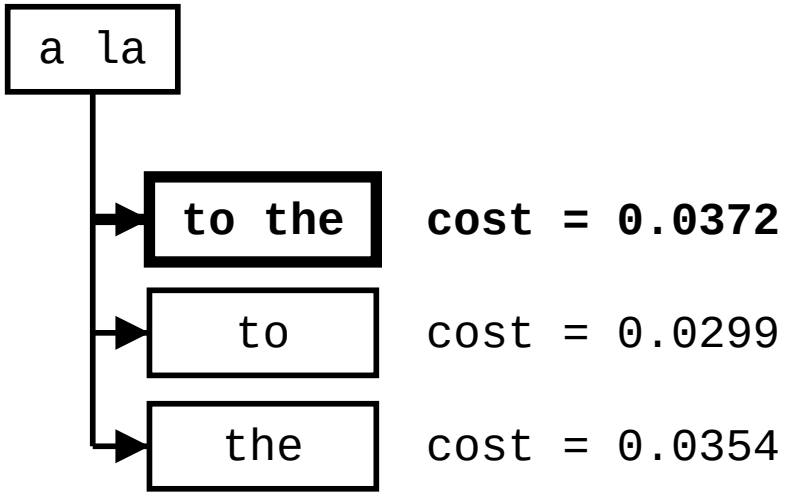


- *Estimate cost* to translate remaining part of input
- Step 1: estimate future cost for each *translation option*
  - look up translation model cost
  - estimate language model cost (no prior context)
  - ignore reordering model cost

  → LM * TM = p(to) * p(to the|a la)

JOHNS HOPKINS UNIVERSITY

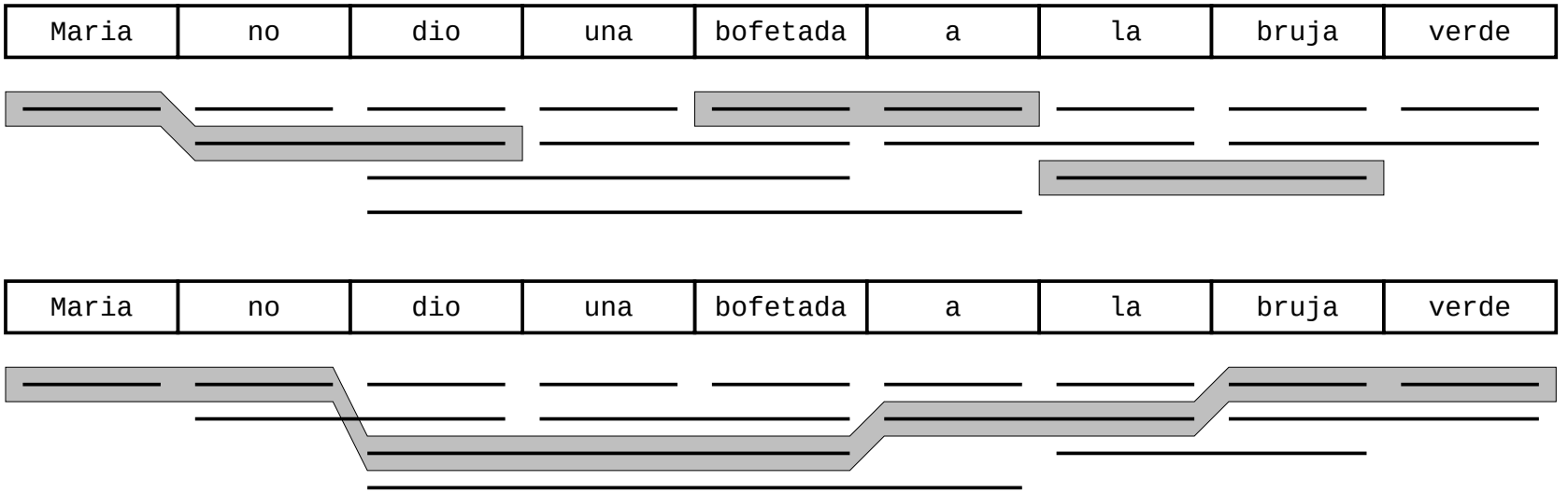
Future Cost Estimation: Step 2



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

JOHNS HOPKINS UNIVERSITY

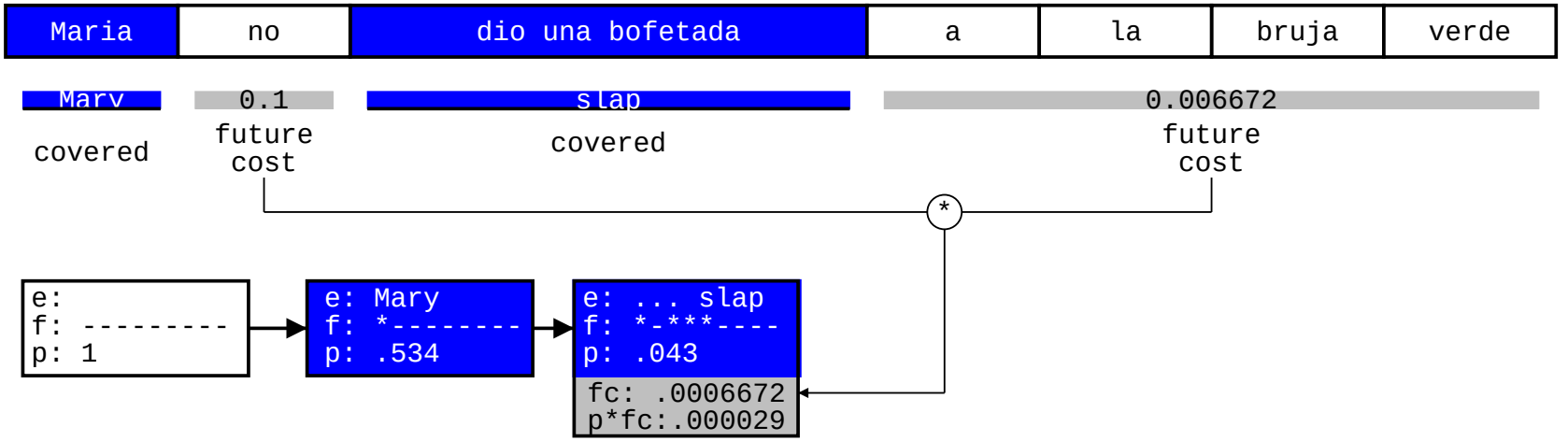
Future Cost Estimation: Step 3



- 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*

JOHNS HOPKINS UNIVERSITY

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

JOHNS HOPKINS UNIVERSITY

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

JOHNS HOPKINS UNIVERSITY

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

JOHNS HOPKINS UNIVERSITY

Automatic evaluation

- Why **automatic evaluation** metrics?
  - Manual evaluation is *too slow*
  - Evaluation on large test sets *reveals minor improvements*
  - **Automatic tuning** to improve machine translation performance
- History
  - Word Error Rate
  - **BLEU** since 2002
- BLEU in short: *Overlap with reference* translations

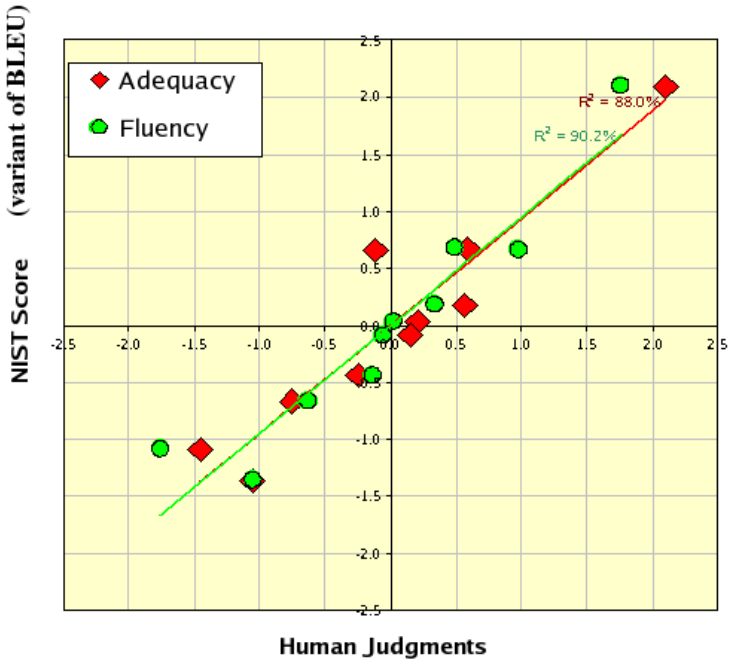
JOHNS HOPKINS UNIVERSITY

Automatic evaluation

- Reference Translation
  - the gunman was shot to death by the police .
- System Translations
  - **the gunman was police kill .**
  - **wounded police jaya of**
  - **the gunman was shot dead by the police .**
  - **the gunman arrested by police kill .**
  - **the gunmen were killed .**
  - **the gunman was shot to death by the police .**
  - **gunmen were killed by police ?SUB>0 ?SUB>0**
  - **al by the police .**
  - **the ringer is killed by the police .**
  - **police killed the gunman .**
- Matches
  - green = 4 gram match (good!)
  - red = word not matched (bad!)

JOHNS HOPKINS UNIVERSITY

Automatic evaluation

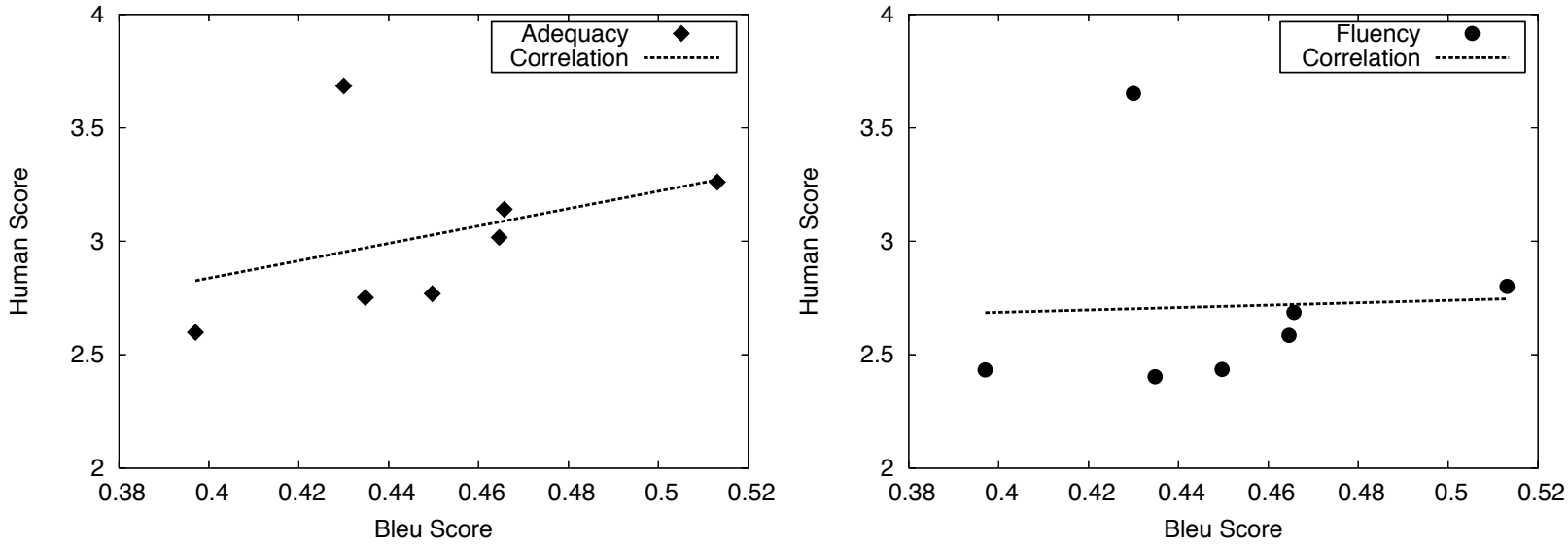


- BLEU **correlates** with human judgement
  - **multiple reference translations** may be used

[from George Doddington, NIST]

JOHNS HOPKINS UNIVERSITY

Correlation? [Callison-Burch et al., 2006]



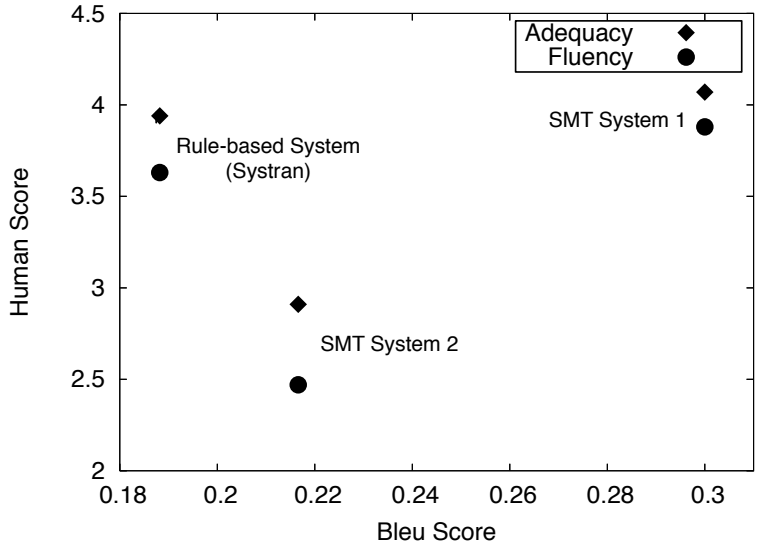
[from Callison-Burch et al., 2006, EACL]

- DARPA/NIST MT Eval 2005
  - Mostly statistical systems (all but one in graphs)
  - One submission **manual post-edit** of statistical system's output

  → Good adequacy/fluency scores *not reflected* by BLEU

JOHNS HOPKINS UNIVERSITY

Correlation? [Callison-Burch et al., 2006]



[from Callison-Burch et al., 2006, EACL]

- Comparison of
  - *good statistical* system: high BLEU, high adequacy/fluency
  - *bad statistical* sys. (trained on less data): low BLEU, low adequacy/fluency
  - *Systran*: lowest BLEU score, but high adequacy/fluency

JOHNS HOPKINS UNIVERSITY

Automatic evaluation: outlook

- Research questions
  - why does BLEU *fail* Systran and manual post-edits?
  - how can this *overcome* with novel evaluation metrics?
- Future of automatic methods
  - automatic metrics too *useful* to be abandoned
  - evidence still supports that during *system development*, a better BLEU indicates a better system
  - *final assessment* has to be human judgement

JOHNS HOPKINS UNIVERSITY

Output of Chinese-English system

In the First Two Months Guangdong's Export of High-Tech Products 3.76 Billion US Dollars

Xinhua News Agency, Guangzhou, March 16 (Reporter Chen Jizhong) - The latest statistics show that between January and February this year, Guangdong's export of high-tech products 3.76 billion US dollars, with a growth of 34.8% and accounted for the province's total export value of 25.5%. The export of high-tech products bright spots frequently now, the Guangdong provincial foreign trade and economic growth has made important contributions. Last year, Guangdong's export of high-tech products 22.294 billion US dollars, with a growth of 31 percent, an increase higher than the province's total export growth rate of 27.2 percent; exports of high-tech products net increase 5.270 billion us dollars, up for the traditional labor-intensive products as a result of prices to drop from the value of domestic exports decreased.

In the Suicide explosion in Jerusalem

Xinhua News Agency, Jerusalem, March 17 (Reporter bell tsui flower nie Xiaoyang) - A man on the afternoon of 17 in Jerusalem in the northern part of the residents of rammed a bus near ignition of carry bomb, the wrongdoers in red-handed was killed and another nine people were slightly injured and sent to hospital for medical treatment.

JOHNS HOPKINS UNIVERSITY

Partially excellent translations

In the First Two Months Guangdong's Export of High-Tech Products 3.76 Billion US Dollars

Xinhua News Agency, Guangzhou, March 16 (Reporter Chen Jizhong) - The latest statistics show that between January and February this year, Guangdong's export of high-tech products 3.76 billion US dollars, with a growth of 34.8% and accounted for the province's total export value of 25.5%. The export of high-tech products bright spots frequently now, the Guangdong provincial foreign trade and economic growth has made important contributions. Last year, Guangdong's export of high-tech products 22.294 billion US dollars, with a growth of 31 percent, an increase higher than the province's total export growth rate of 27.2 percent; exports of high-tech products net increase 5.270 billion US dollars, up for the traditional labor-intensive products as a result of prices to drop from the value of domestic exports decreased.

In the Suicide explosion in Jerusalem

Xinhua News Agency, Jerusalem, March 17 (Reporter bell tsui flower nie Xiaoyang) - A man on the afternoon of 17 in Jerusalem in the northern part of the residents of rammed a bus near ignition of carry bomb, the wrongdoers in red-handed was killed and another nine people were slightly injured and sent to hospital for medical treatment.

JOHNS HOPKINS UNIVERSITY

Mangled grammar

In the First Two Months Guangdong's Export of High-Tech Products 3.76 Billion US Dollars

Xinhua News Agency, Guangzhou, March 16 (Reporter Chen Jizhong) - The latest statistics show that between January and February this year, Guangdong's export of high-tech products 3.76 billion US dollars, with a growth of 34.8% and accounted for the province's total export value of 25.5%. The export of high-tech products bright spots frequently now, the Guangdong provincial foreign trade and economic growth has made important contributions. Last year, Guangdong's export of high-tech products 22.294 billion US dollars, with a growth of 31 percent, an increase higher than the province's total export growth rate of 27.2 percent; exports of high-tech products net increase 5.270 billion us dollars, up for the traditional labor-intensive products as a result of prices to drop from the value of domestic exports decreased.

In the Suicide explosion in Jerusalem

Xinhua News Agency, Jerusalem, March 17 (Reporter bell tsui flower nie Xiaoyang) - A man on the afternoon of 17 in Jerusalem in the northern part of the residents of rammed a bus near ignition of carry bomb, the wrongdoers in red-handed was killed and another nine people were slightly injured and sent to hospital for medical treatment.