Course Wrap Up

Machine Translation
Lecture 26

Instructor: Chris Callison-Burch
TAs: Mitchell Stern, Justin Chiu

Website: mt-class.org/penn

Course co-inventors

Adam Lopez
Edinburgh

Matt Post
JHU

Chris Dyer
CMU

Statistical Machine Translation

Develop a statistical model of translation that can be learned from data and used to predict the correct English translation of new foreign sentences.

Lexical Translation Models

$$p(e_i \mid f_{a_i})$$

$$\hat{p}_{\text{MLE}}(e \mid \texttt{Haus}) = \begin{cases} 0.696 & \text{if } e = \texttt{house} \\ 0.279 & \text{if } e = \texttt{home} \\ 0.014 & \text{if } e = \texttt{shell} \\ 0.011 & \text{if } e = \texttt{household} \\ 0 & \text{otherwise} \end{cases}$$

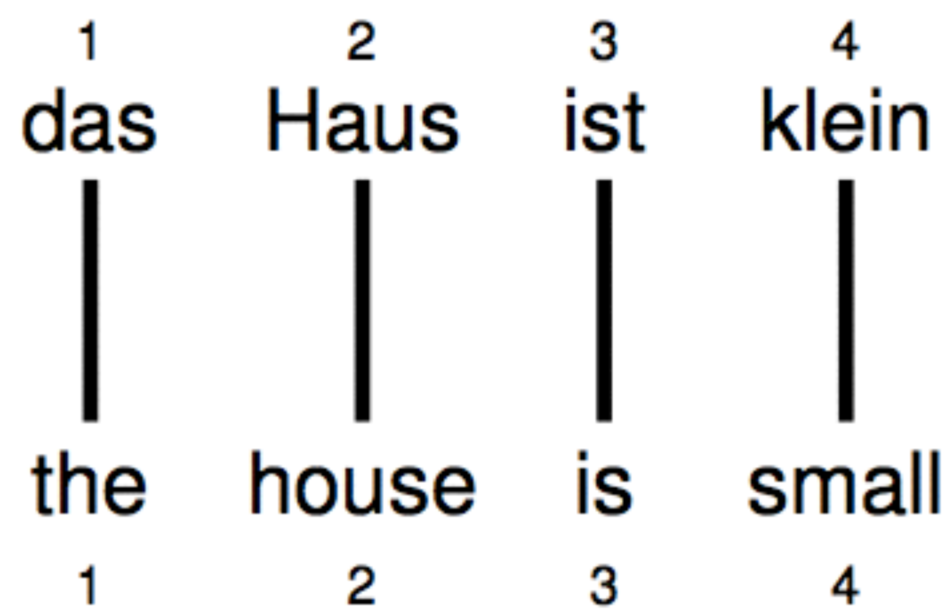
Lexical Translation Models

$$p(\mathbf{e} \mid \mathbf{f}, m) = \sum_{\mathbf{a} \in [0,n]^m} p(\mathbf{a} \mid \mathbf{f}, m) \times \prod_{i=1}^{m} p(e_i \mid f_{a_i})$$

Alignment × Translation | Alignment

Alignment

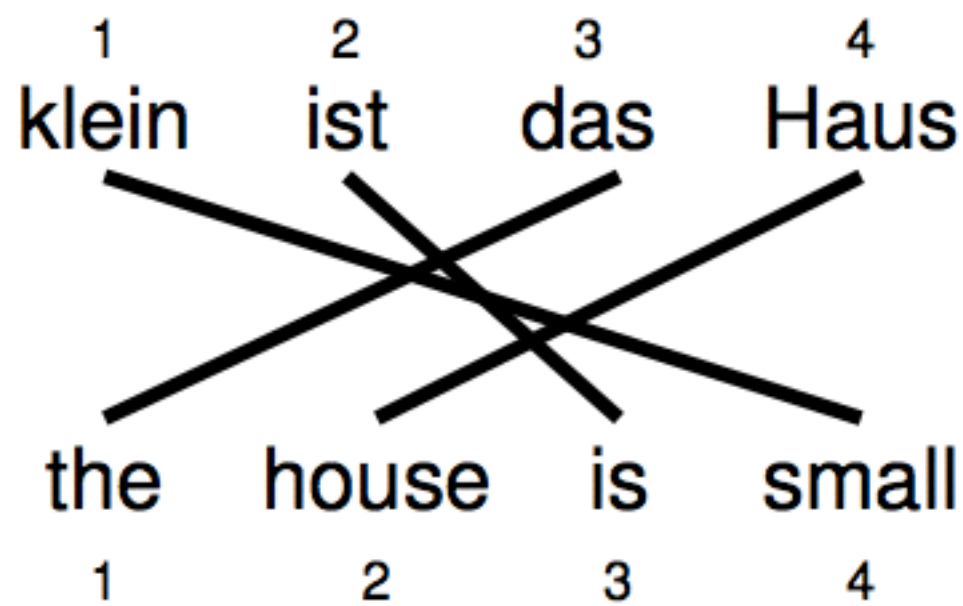
- Alignments can be visualized in by drawing links between two sentences, and they are represented as vectors of positions:



$$\mathbf{a} = (1, 2, 3, 4)^{\top}$$

Reordering

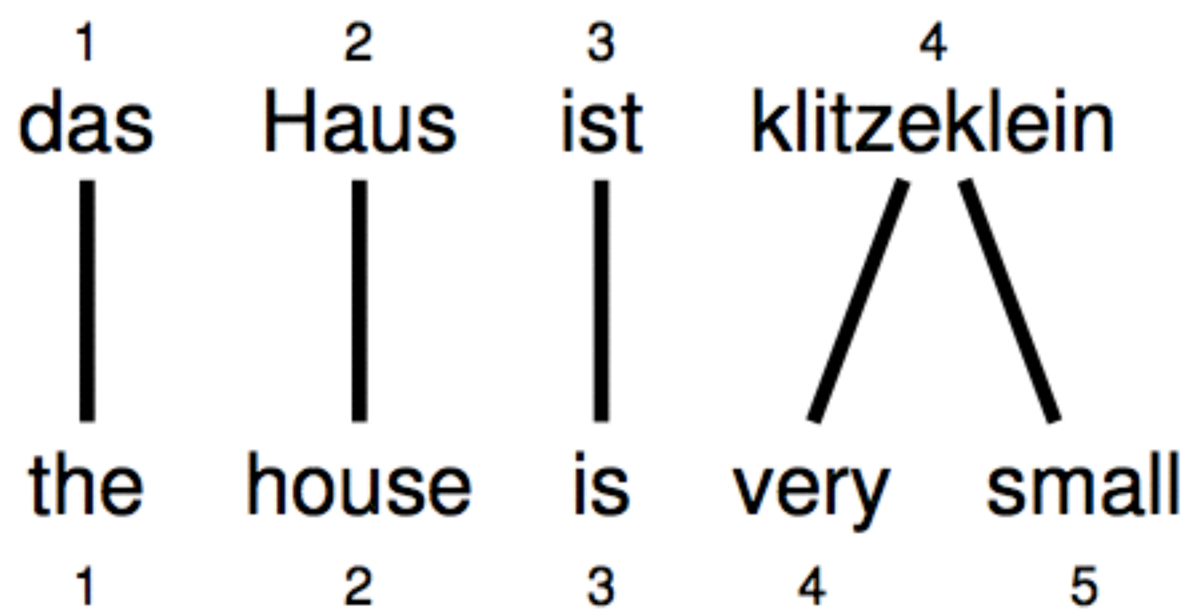
- Words may be reordered during translation.



$$\mathbf{a} = (3, 4, 2, 1)^\top$$

One-to-many Translation

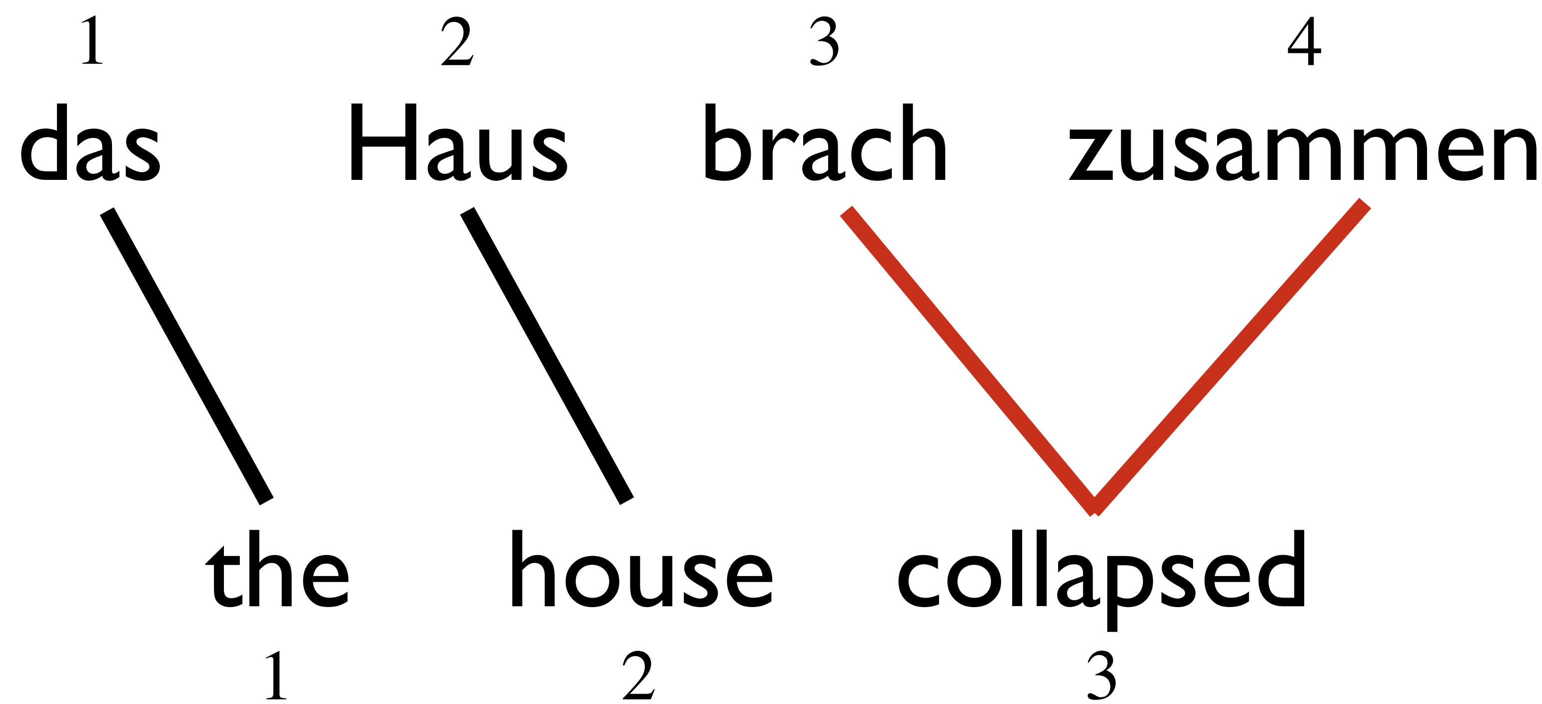
- A source word may translate into **more than one** target word



$$\mathbf{a} = (1, 2, 3, 4, 4)^\top$$

Many-to-one Translation

- More than one source word may not translate as a unit in lexical translation



$\mathbf{a} = ???$ $\qquad$ $\mathbf{a} = (1, 2, (3, 4)^\top)^\top$?

IBM Model 1

- Simplest possible lexical translation model
- Additional assumptions
  - The m alignment decisions are independent
  - The alignment distribution for each a_i is uniform over all source words and NULL

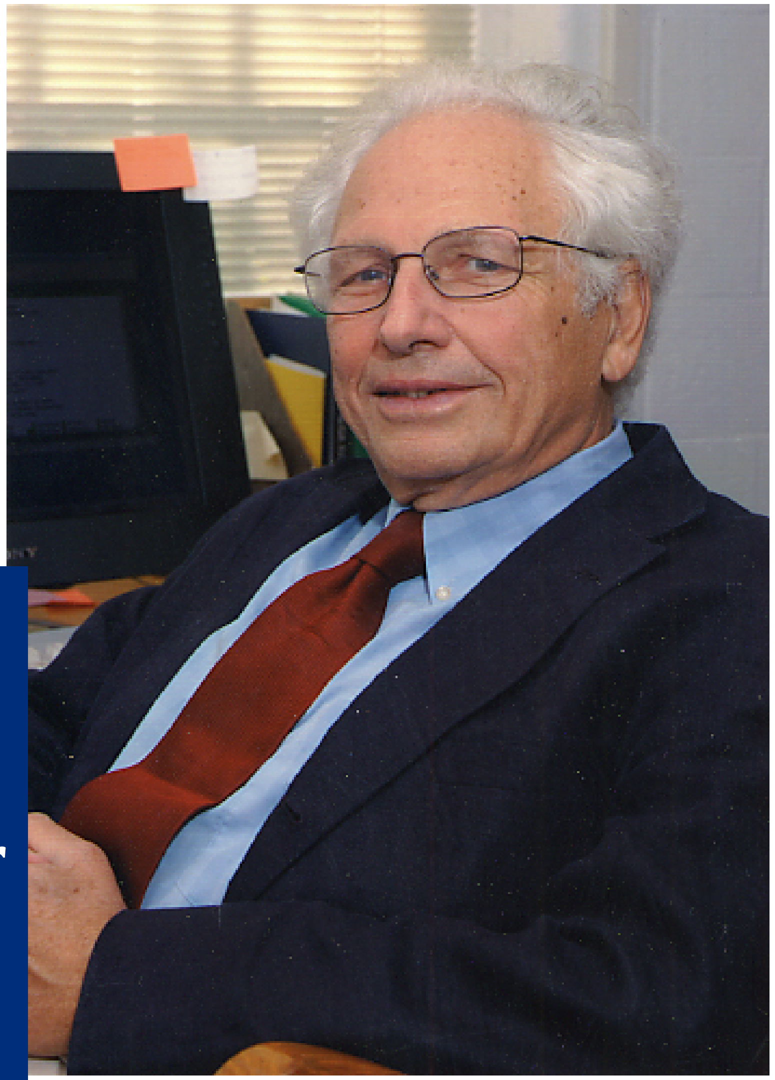
$$
\begin{aligned}
&\text{for each } i \in [1, 2, \ldots, m] \\
&\qquad\qquad a_i \sim \text{Uniform}(0, 1, 2, \ldots, n) \\
&\qquad\qquad e_i \sim \text{Categorical}(\boldsymbol{\theta}_{f_{a_i}})
\end{aligned}
$$

Historical Note

IBM Models

Renaissance

"The validity of a statistical (information theoretic) approach to MT has indeed been recognized, as the authors mention, by Weaver as early as 1949. And was universally recognized as mistaken by 1950 (cf. Hutchins, MT – Past, Present, Future, Ellis Horwood, 1986, p. 30ff and references therein). The crude force of computers is not science. The paper is simply beyond the scope of COLING."



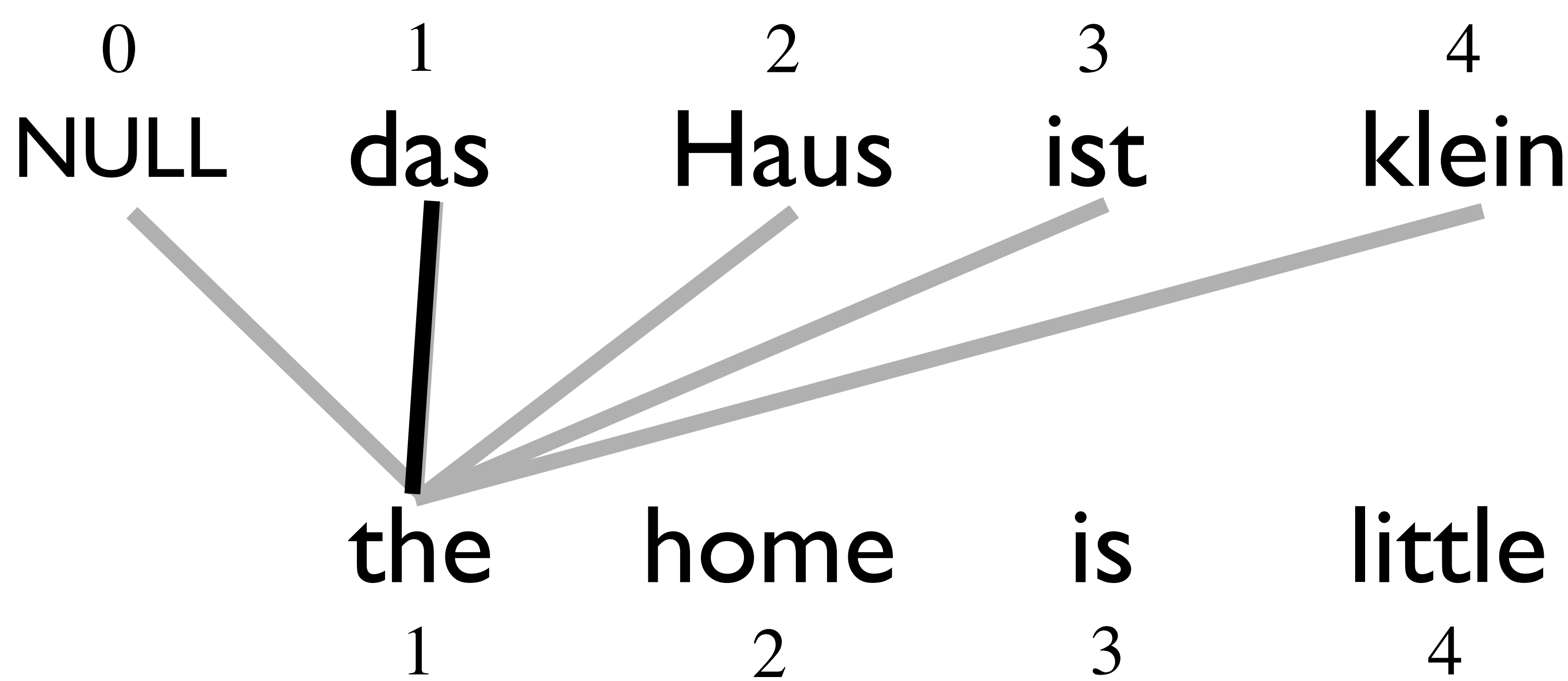
Fred Jelinek
(1932-2010)

The Center For Language and Speech Processing
at the Johns Hopkins University

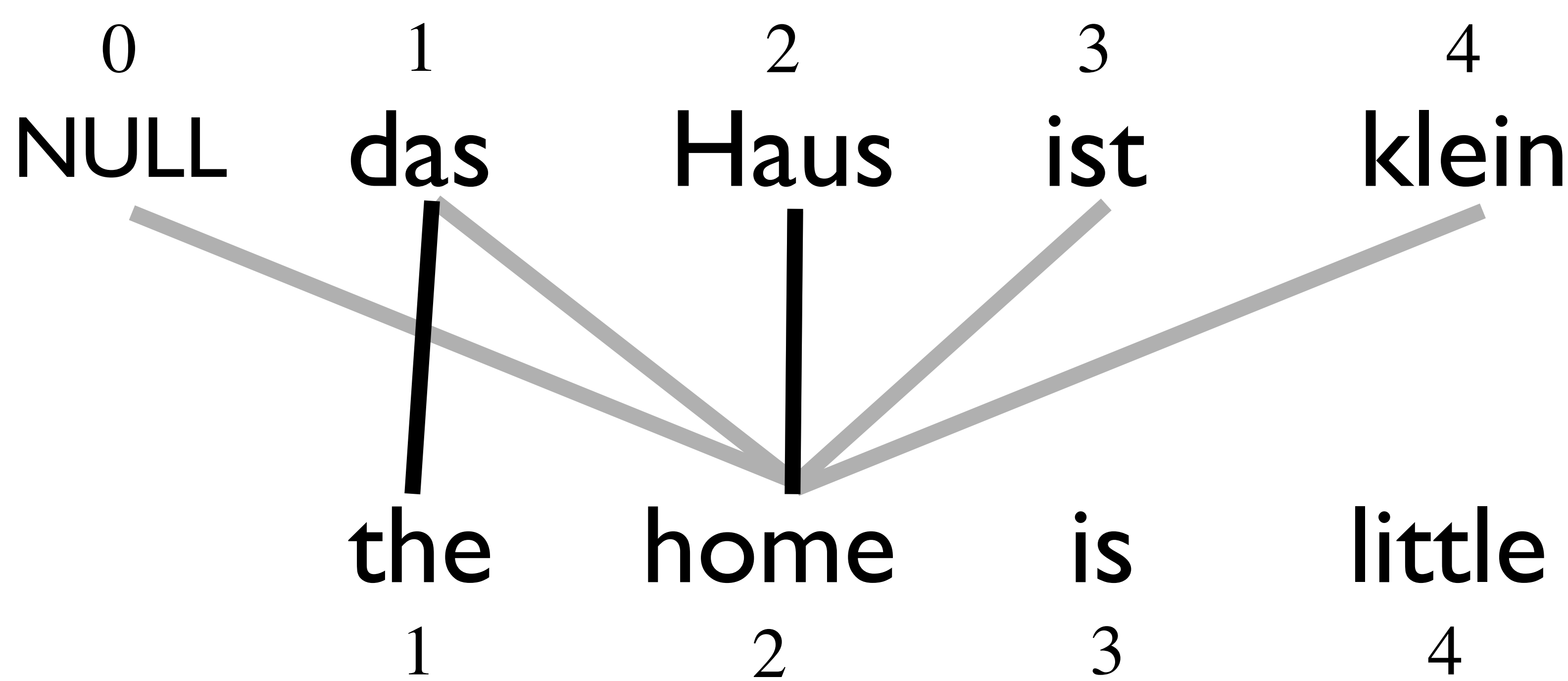
1. Introduction

The growing availability of bilingual, machine-readable texts has stimulated interest

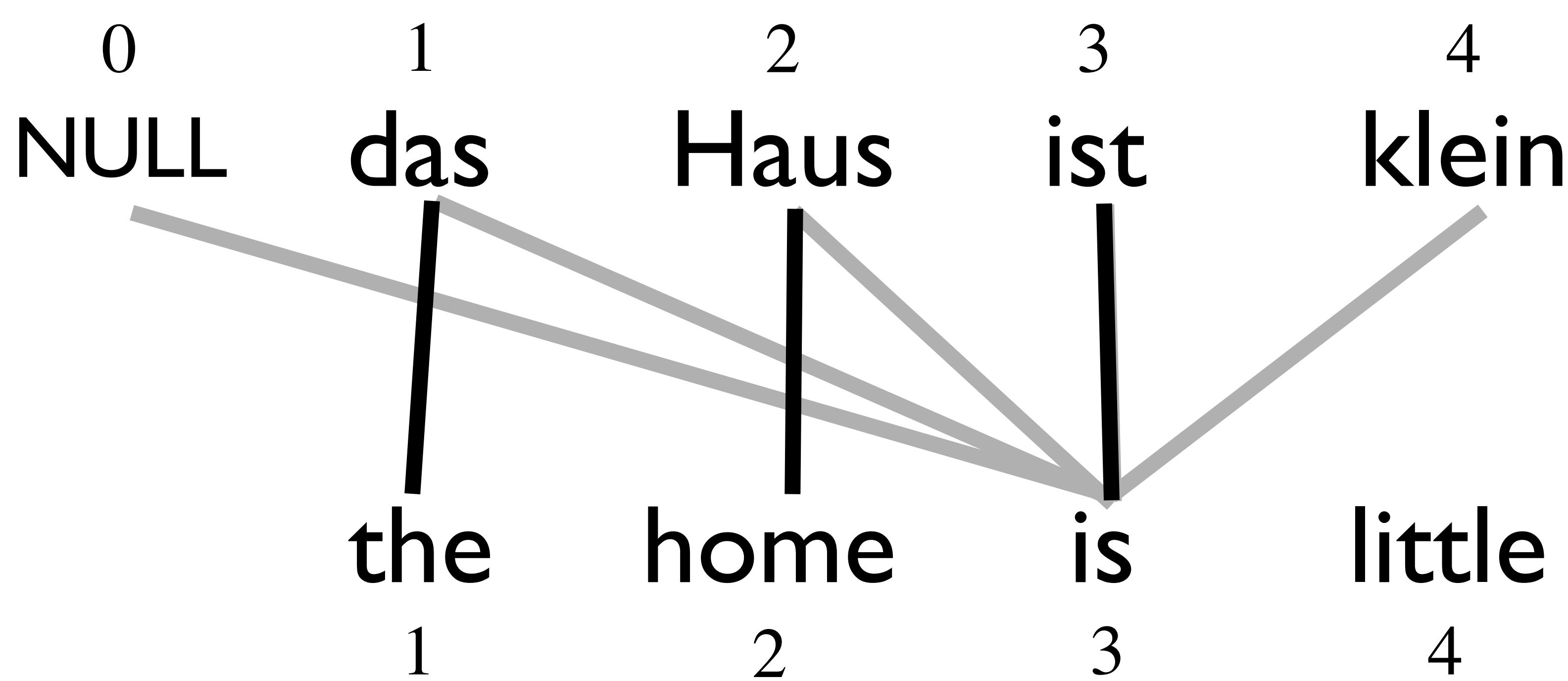
Finding the Viterbi Alignment



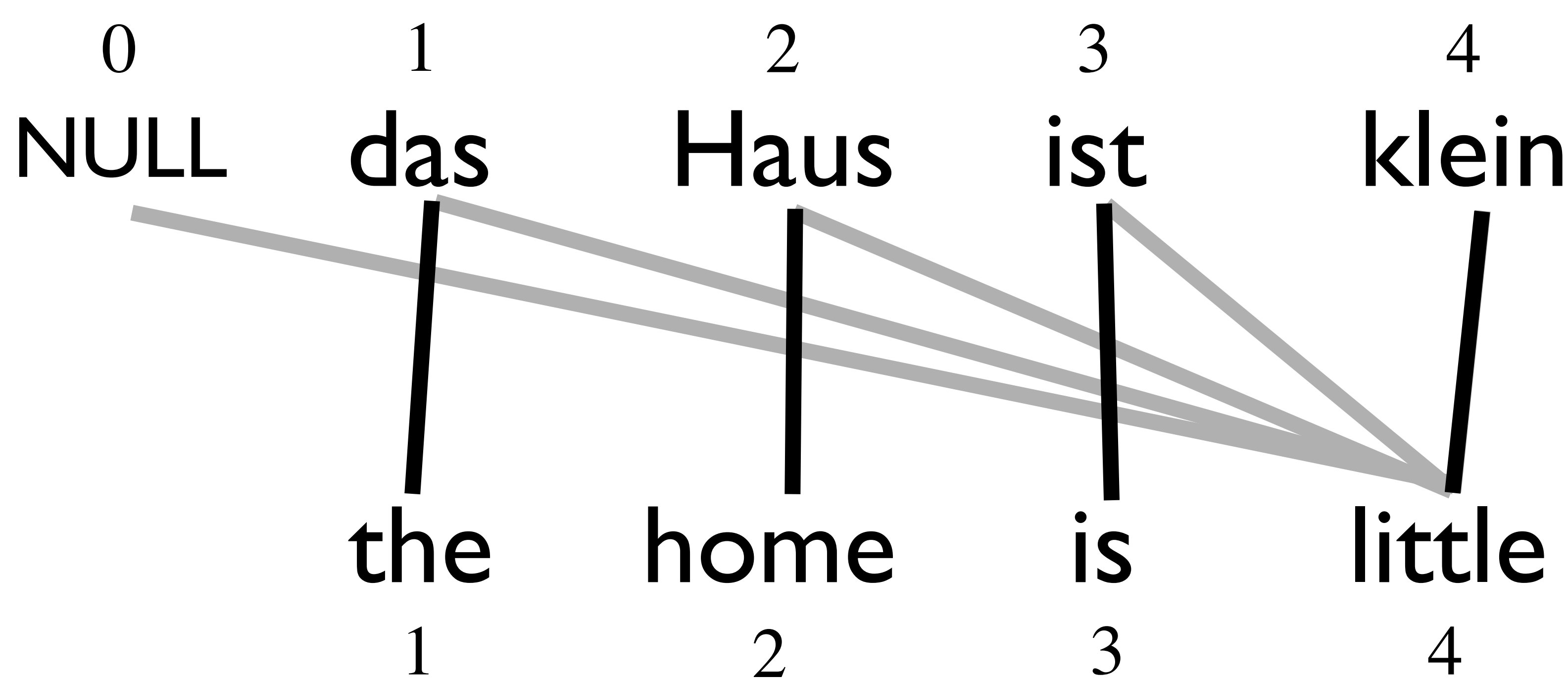
Finding the Viterbi Alignment



Finding the Viterbi Alignment

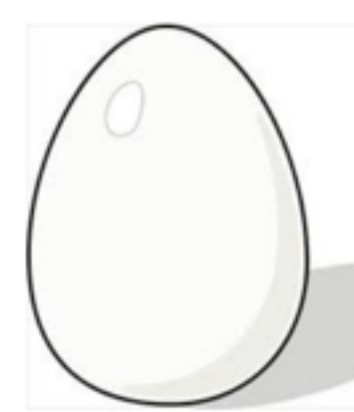


Finding the Viterbi Alignment



Learning Lexical Translation Models

- How do we learn the parameters $p(e \mid f)$
- "Chicken and egg" problem
  - If we had the alignments, we could estimate the parameters (MLE)
  - If we had parameters, we could find the most likely alignments



You implemented your own word aligner

Phrase-based Translation

- What are the atomic units?
  - Lexical translation: **words**
  - Phrase-based translation: **phrases**
- Benefits
  - many-to-many translation
  - use of local context in translation
- **Standard model used by Google, Microsoft ...**

Phrase-based Translation

- With a **latent variable**, we introduce a decomposition into **phrases** which translate **independently**:

$$p(\mathbf{f}, \mathbf{a} \mid \mathbf{e}) = p(\mathbf{a}) \prod_{\langle \bar{\mathbf{e}}, \bar{\mathbf{f}} \rangle \in \mathbf{a}} p(\bar{\mathbf{f}} \mid \bar{\mathbf{e}})$$

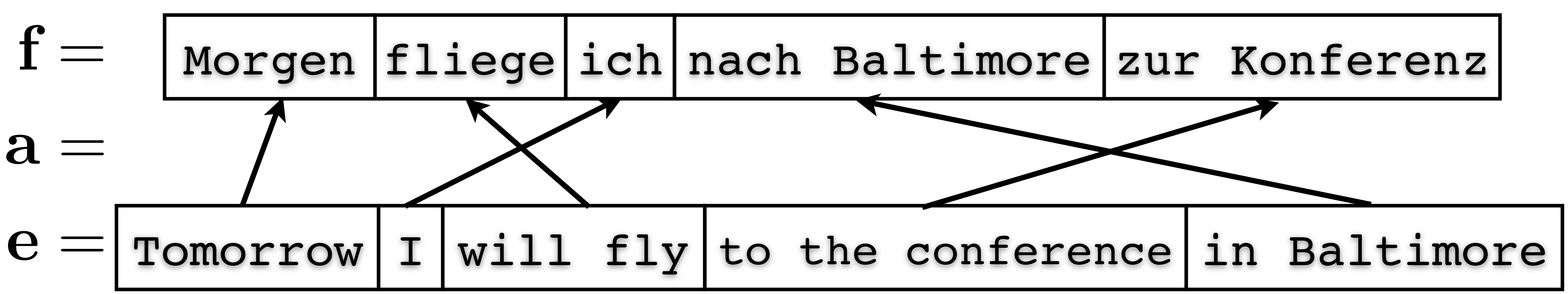
$\mathbf{f} =$ `Morgen fliege ich nach Baltimore zur Konferenz`

$\mathbf{e} =$ `Tomorrow I will fly to the conference in Baltimore`

Phrase-based Translation

- With a **latent variable**, we introduce a decomposition into **phrases** which translate **independently**:

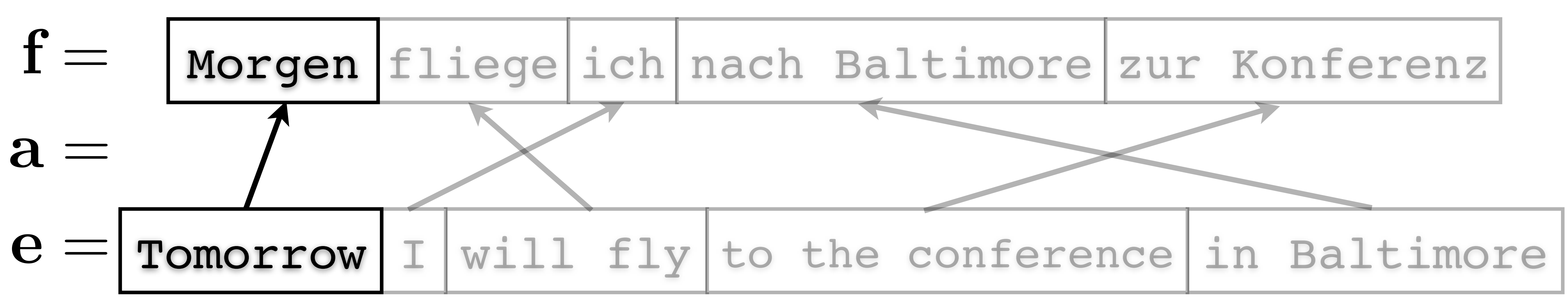
$$p(\mathbf{f}, \mathbf{a} \mid \mathbf{e}) = p(\mathbf{a}) \prod_{\langle \bar{\mathbf{e}}, \bar{\mathbf{f}} \rangle \in \mathbf{a}} p(\bar{\mathbf{f}} \mid \bar{\mathbf{e}})$$



Phrase-based Translation

- With a **latent variable**, we introduce a decomposition into **phrases** which translate **independently**:

$$p(\mathbf{f}, \mathbf{a} \mid \mathbf{e}) = p(\mathbf{a}) \prod_{\langle \bar{\mathbf{e}}, \bar{\mathbf{f}} \rangle \in \mathbf{a}} p(\bar{\mathbf{f}} \mid \bar{\mathbf{e}})$$

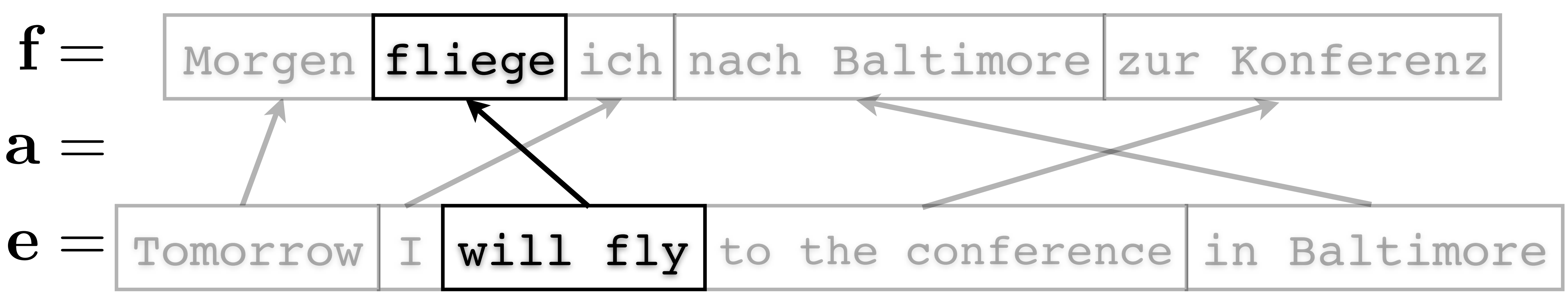


p(Morgen|Tomorrow)

Phrase-based Translation

- With a **latent variable**, we introduce a decomposition into **phrases** which translate **independently**:

$$p(\mathbf{f}, \mathbf{a} \mid \mathbf{e}) = p(\mathbf{a}) \prod_{\langle \bar{\mathbf{e}}, \bar{\mathbf{f}} \rangle \in \mathbf{a}} p(\bar{\mathbf{f}} \mid \bar{\mathbf{e}})$$

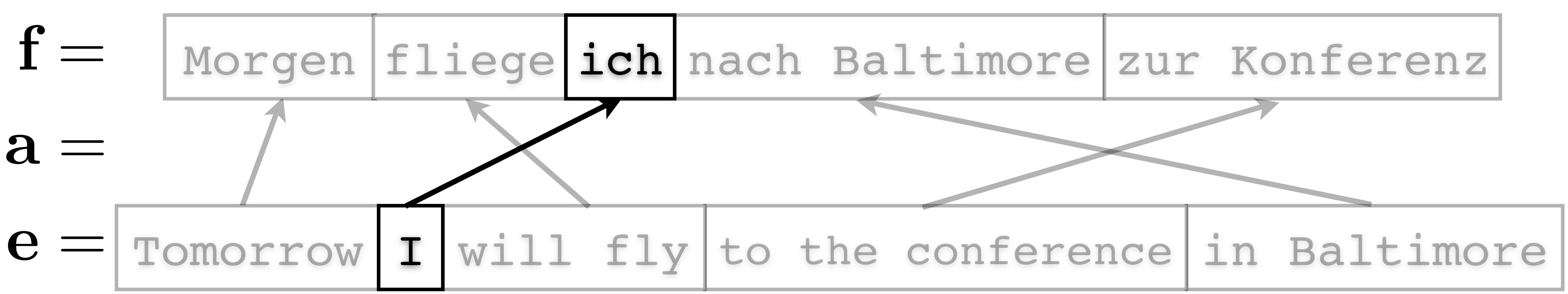


p(Morgen|Tomorrow) x p(fliege|will fly)

Phrase-based Translation

- With a **latent variable**, we introduce a decomposition into **phrases** which translate **independently**:

$$p(\mathbf{f}, \mathbf{a} \mid \mathbf{e}) = p(\mathbf{a}) \prod_{\langle \bar{\mathbf{e}}, \bar{\mathbf{f}} \rangle \in \mathbf{a}} p(\bar{\mathbf{f}} \mid \bar{\mathbf{e}})$$

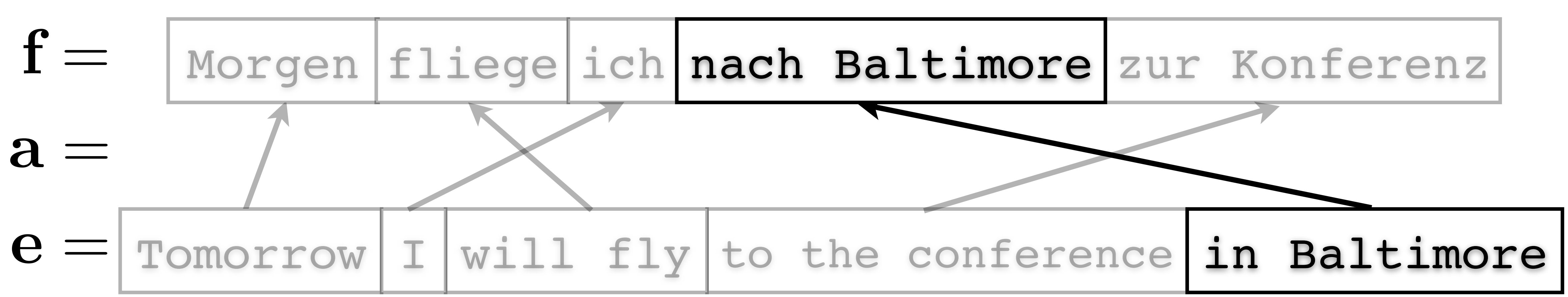


p(Morgen|Tomorrow) x p(fliege|will fly) x p(ich|I)

Phrase-based Translation

- With a **latent variable**, we introduce a decomposition into **phrases** which translate **independently**:

$$p(\mathbf{f}, \mathbf{a} \mid \mathbf{e}) = p(\mathbf{a}) \prod_{\langle \bar{\mathbf{e}}, \bar{\mathbf{f}} \rangle \in \mathbf{a}} p(\bar{\mathbf{f}} \mid \bar{\mathbf{e}})$$



p(Morgen|Tomorrow) x *p*(fliege|will fly) x *p*(ich|I) x ...

Phrase-based Translation

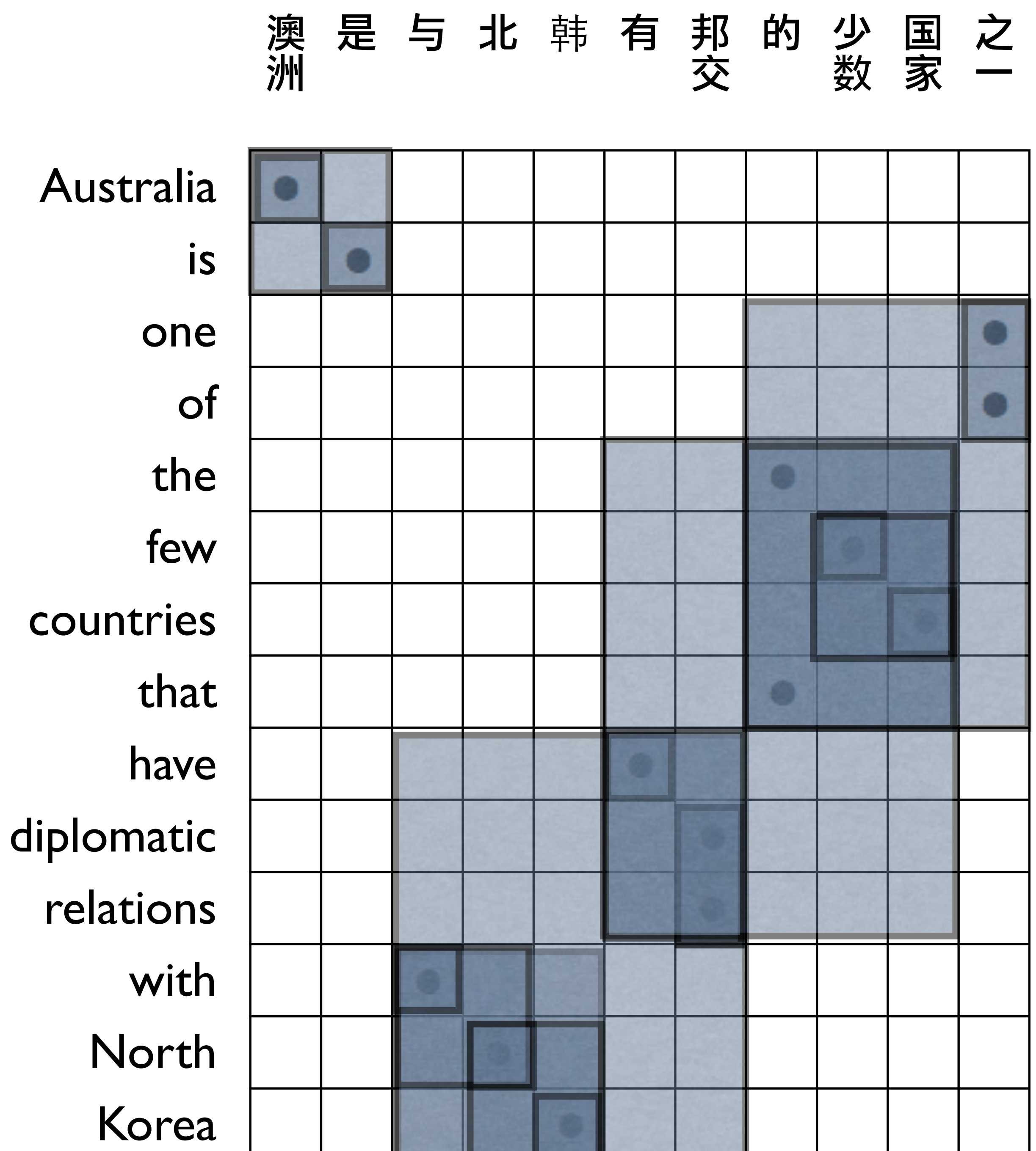
- With a **latent variable**, we introduce a decomposition into **phrases** which translate **independently**:

$$p(\mathbf{f}, \mathbf{a} \mid \mathbf{e}) = p(\mathbf{a}) \prod_{\langle \bar{\mathbf{e}}, \bar{\mathbf{f}} \rangle \in \mathbf{a}} p(\bar{\mathbf{f}} \mid \bar{\mathbf{e}})$$

Marginalize to get p(f|e):

$$p(\mathbf{f} \mid \mathbf{e}) = \sum_{\mathbf{a} \in \mathcal{A}} p(\mathbf{a}) \prod_{\langle \bar{\mathbf{e}}, \bar{\mathbf{f}} \rangle \in \mathbf{a}} p(\bar{\mathbf{f}} \mid \bar{\mathbf{e}})$$

Extracting phrase pairs



澳洲，Australia
是，is
之一，one of
少数，few
国家，countries
有，have
邦交，diplomatic relations
与，with
北，North
韩，Korea

澳洲是，Australia is
少数 国家，few countries
有邦交，have diplomatic relations
与北，with North
北韩，North Korea

的少数 国家，the few countries that
与北韩，with North Korea

之一的少数 国家，one of the the few countries that
与北韩 有邦交，have diplomatic relations with North Korea

有邦交 的少数 国家，the few countries that have diplomatic relations

Phrase Tables

$\bar{\mathbf{f}}$	$\bar{\mathbf{e}}$	$p(\bar{\mathbf{f}} \mid \bar{\mathbf{e}})$
das Thema	the issue	0.41
	the point	0.72
	the subject	0.47
	the thema	0.99
es gibt	there is	0.96
	there are	0.72
morgen	tomorrow	0.9
fliege ich	will I fly	0.63
	will fly	0.17
	I will fly	0.13

Phrases

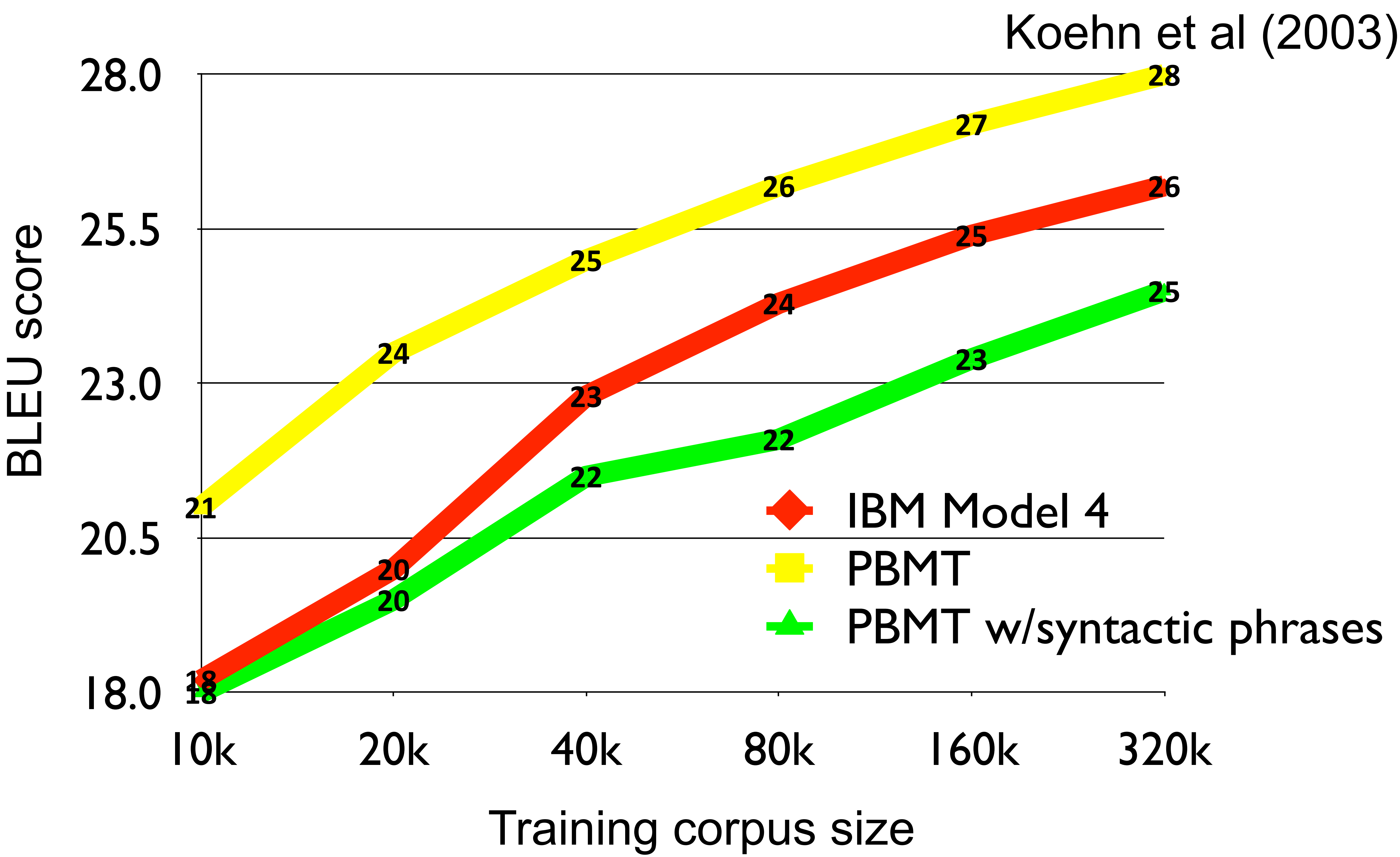
- Contiguous strings of words
- Phrases are not necessarily syntactic constituents
- Usually have maximum limits
- Phrases subsume words (individual words are phrases of length 1)

Linguistic Phrases

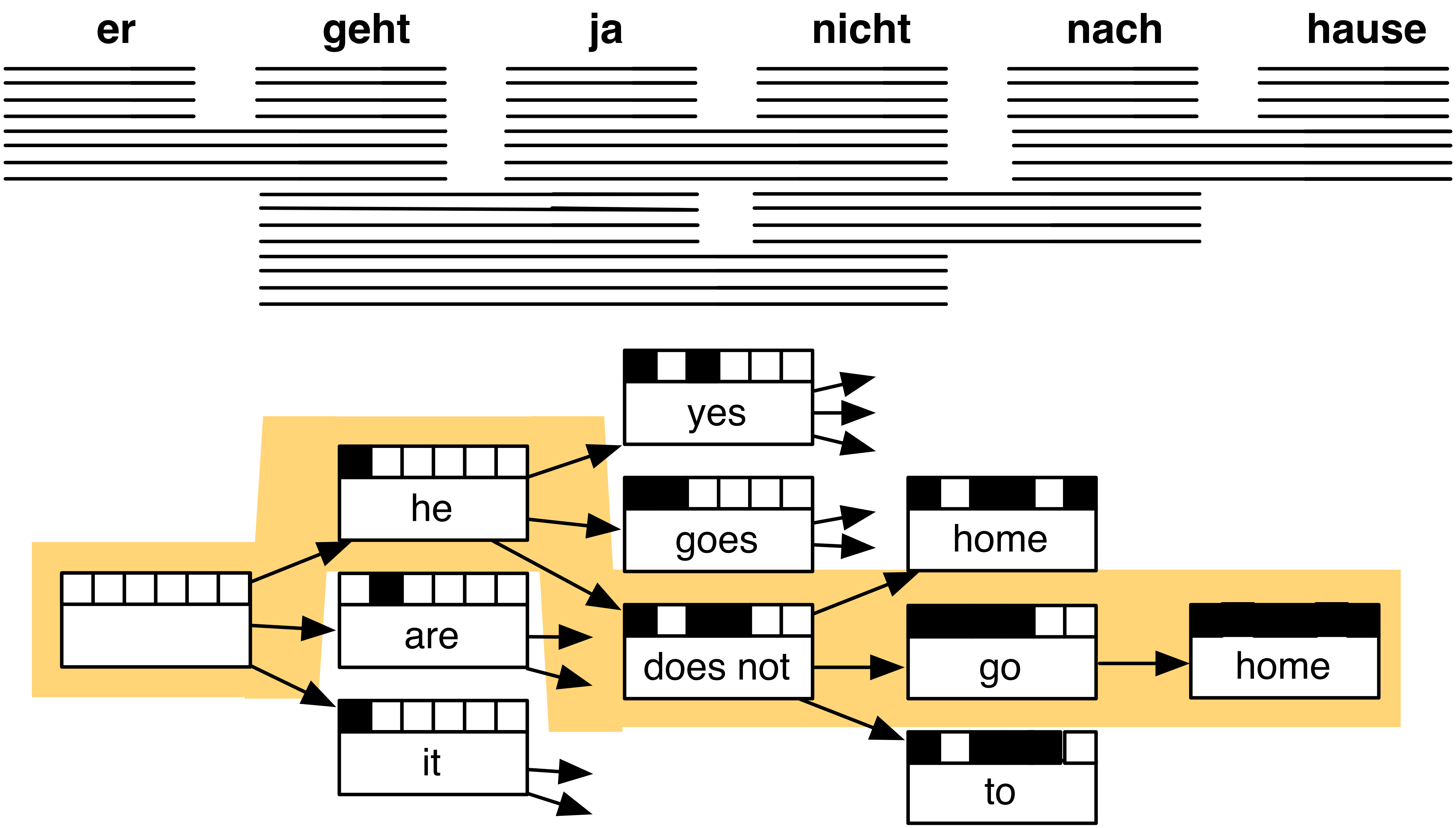
- Model is not limited to linguistic phrases (NPs, VPs, PPs, CPs...)
- Non-constituent phrases are useful

es gibt there is | there are

Syntax hurts

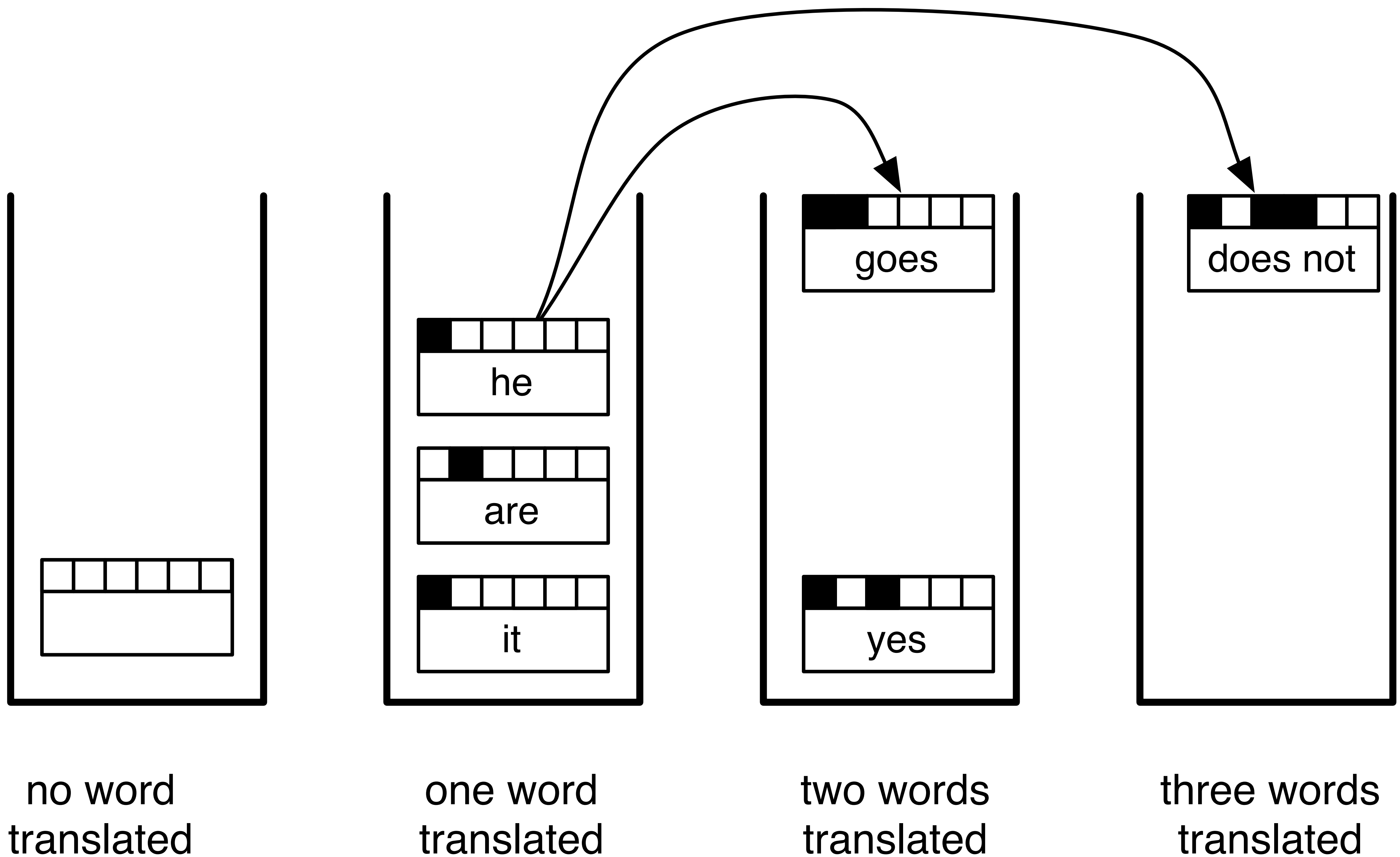


Decoding



er
geht
ja
nicht
nach
hause
yes
he
goes
home
are
does not
go
home
it
to

Decoding


goes
does not
he
are
it
yes
no word translated
one word translated
two words translated
three words translated

Decoding

```
place empty hypothesis into stack 0
for all stacks 0...n-1 do
    for all hypotheses in stack do
        for all translation options do
            if applicable then
                create new hypothesis
                place in stack
                recombine with existing hypothesis if possible
                prune stack if too big
            end if
        end for
    end for
end for
```

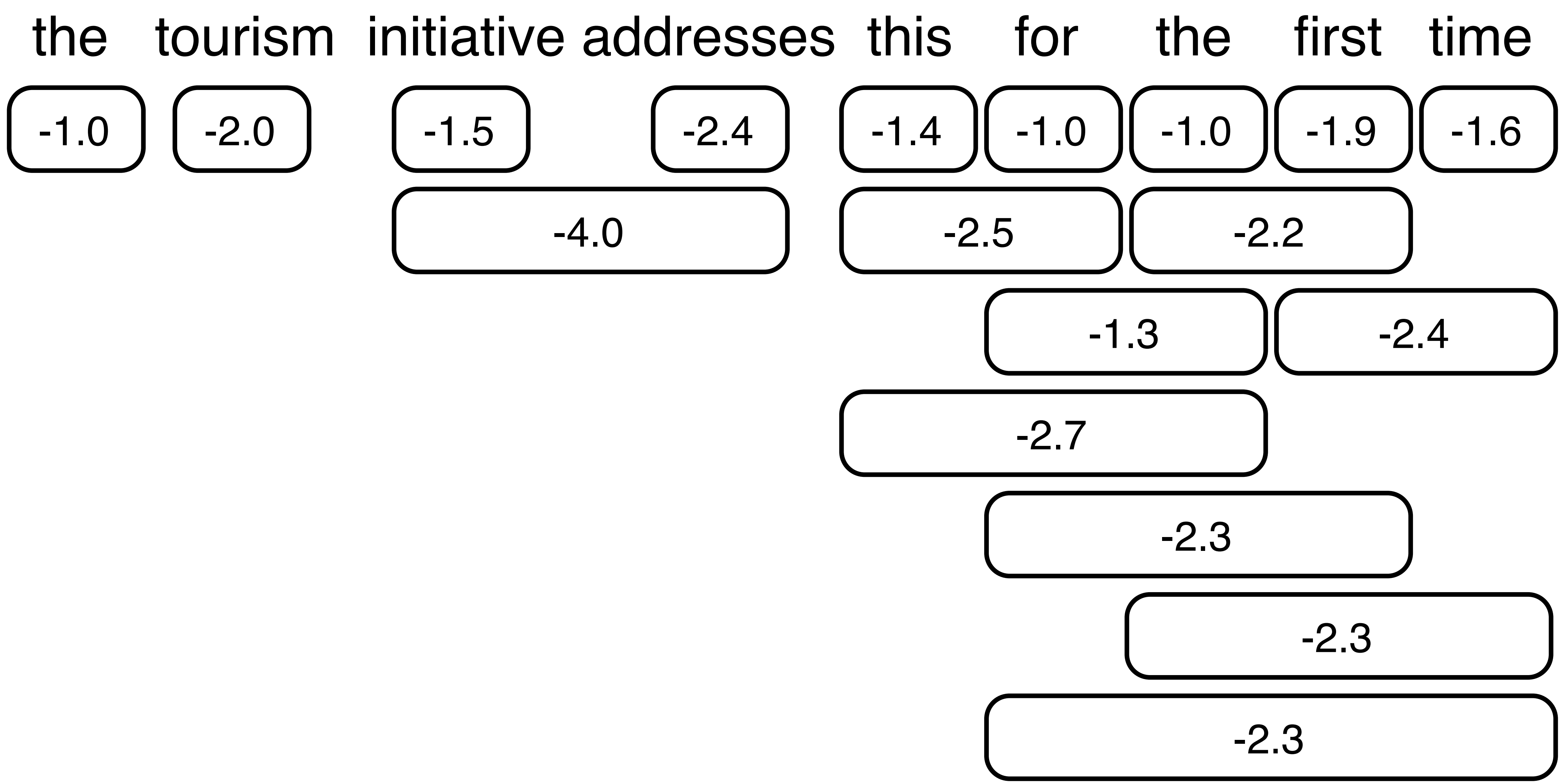
Decoding complexity

- Finding the best hypothesis is NP-hard
  - Even with no language model, there are an exponential number of states!
  - Solution 1: limit reordering
  - Solution 2: (lossy) pruning

Search Errors

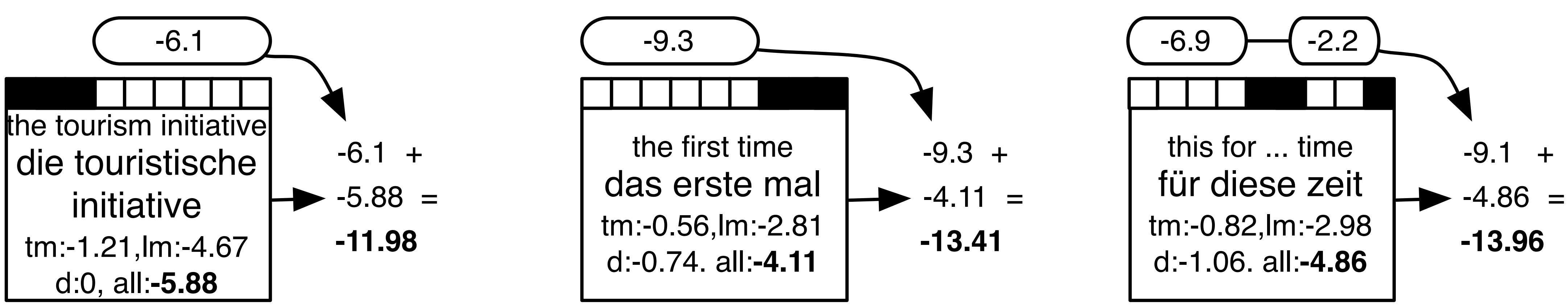
- We are using a **heuristic search** to prune the search space
- There are no guarantees of admissibility (like in A* search)
- We may therefore prune out a partial hypothesis that would have lead to the most probable translation, if we hadn't pruned it early on

Future Cost Estimation



Future Cost Estimation

first word	future cost estimate for n words (from first)								
	1	2	3	4	5	6	7	8	9
the	-1.0	-3.0	-4.5	-6.9	-8.3	-9.3	-9.6	-10.6	-10.6
tourism	-2.0	-3.5	-5.9	-7.3	-8.3	-8.6	-9.6	-9.6	
initiative	-1.5	-3.9	-5.3	-6.3	-6.6	-7.6	-7.6		
addresses	-2.4	-3.8	-4.8	-5.1	-6.1	-6.1			
this	-1.4	-2.4	-2.7	-3.7	-3.7				
for	-1.0	-1.3	-2.3	-2.3					
the	-1.0	-2.2	-2.3						
first	-1.9	-2.4							
time	-1.6								



Linguistic Phrases

- Model is not limited to linguistic phrases (NPs, VPs, PPs, CPs...)
- Non-constituent phrases are useful

es gibt there is | there are

- Is a "good" phrase more likely to be
  [P NP] or [governor P]
  Why? How would you figure this out?

You implemented your own phrase-based decoder

Evaluating Translation Quality

- Why do we want to do it?
- Want to rank systems
- Want to evaluate incremental changes
- What to make scientific claims

Goals for Automatic Evaluation

- No cost evaluation for incremental changes
- Ability to rank systems
- Ability to identify which sentences we're doing poorly on, and categorize errors
- Correlation with human judgments
- Interpretability of the score

Methodology

- Comparison against reference translations
- Intuition: closer we get to human translations, the better we're doing
- Can't use WER like in speech recognition
  - *This shows how easy it is to recognize speech*
  - *This shows how easy it is to wreck a nice beach*

Problems with WER

- In machine translation there can be many possible (and equally valid) ways of translating a sentence
- *This shows how easy it is to recognize speech*
- *It illustrates how simple it is to transcribe the spoken word*
- Clauses can move around
- *This shows that recognizing speech is easy*

BLEU

- <u>B</u>i<u>L</u>ingual <u>E</u>valuation <u>U</u>nderstudy
- Uses multiple reference translations
- Look for n-grams that occur anywhere in the sentence

Multiple References

Ref 1	Orejuela appeared calm as he was led to the American plane which will take him to Miami, Florida.
Ref 2	Orejuela appeared calm while being escorted to the plane that would take him to Miami, Florida.
Ref 3	Orejuela appeared calm as he was being led to the American plane that was to carry him to Miami in Florida.
Ref 4	Orejuela seemed quite calm as he was being led to the American plane that would take him to Miami in Florida.

n-gram precision

$$p_n = \frac{\sum_{S \in C} \sum_{ngram \in S} Count_{matched}(ngram)}{\sum_{S \in C} \sum_{ngram \in S} Count(ngram)}$$

American, Florida, Miami, Orejuela, **appeared**, as, being, calm, carry, escorted, he, him, in, led, **plane**, quite, seemed, take, that, the, **to**, **to**, to, **was** , was, **which**, while, **will**, would, **,, .**

1-gram precision = 15/18

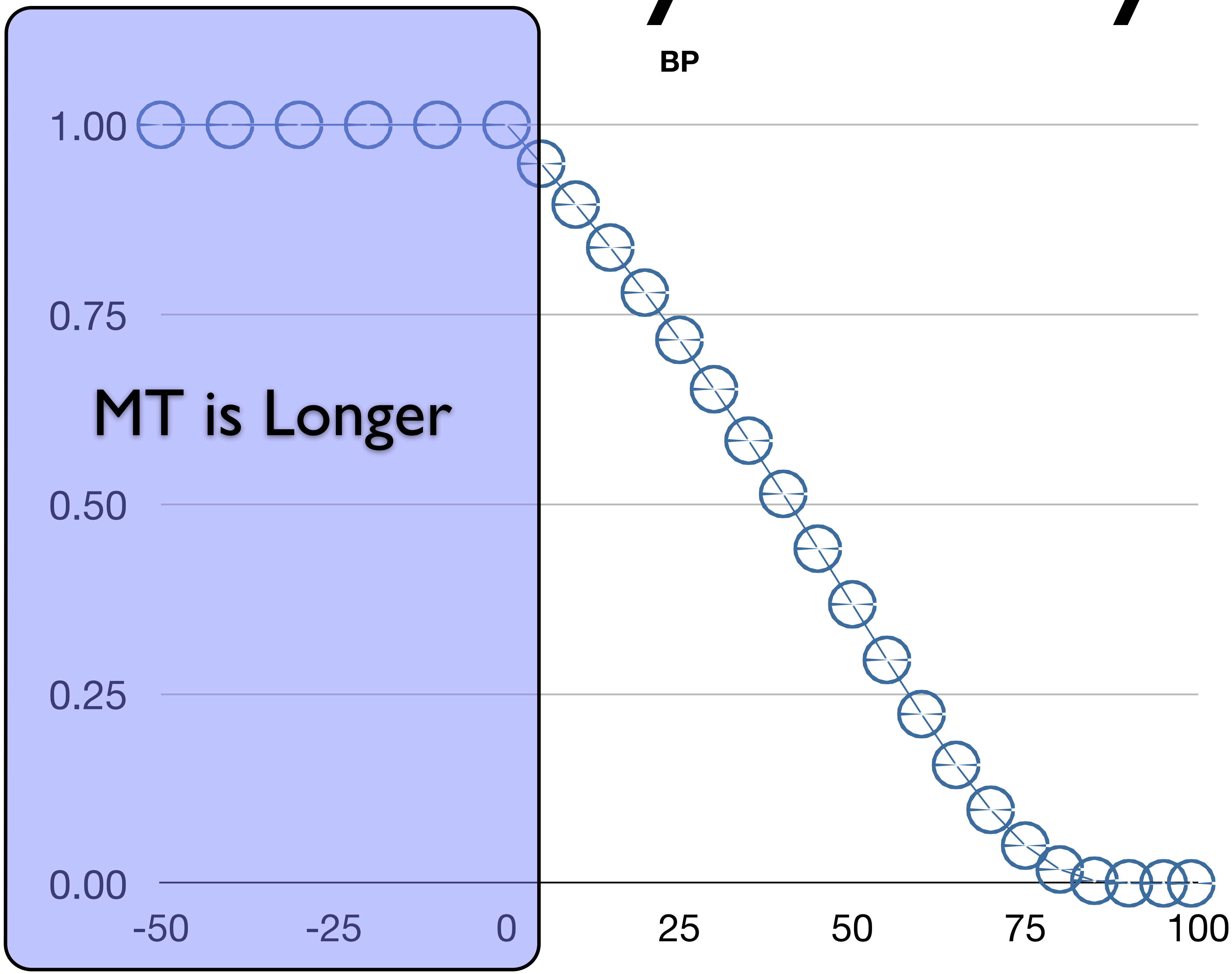
Hyp	**appeared calm** when **he was** taken **to the American plane** , **which will to Miami , Florida .**

American plane, **Florida .**, **Miami ,**, Miami in, Orejuela appeared, Orejuela seemed, **appeared calm**, as he, being escorted, being led, calm as, calm while, carry him, escorted to, **he was**, him to, in Florida, led to, plane that, plane which, quite calm, seemed quite, take him, that was, that would, **the American**, the plane, **to Miami**, to carry, **to the**, was being, was led, was to, **which will**, while being, will take, would take, , Florida

2-gram precision = 10/17

Hyp	**appeared calm** when **he was** taken **to the American plane** , **which will to Miami , Florida .**

Brevity Penalty



Difference with effective reference length (%)

Manual Evaluation

Source: Estos tejidos están analizados, transformados y congelados antes de ser almacenados en Hema-Québec, que gestiona también el único banco público de sangre del cordón umbilical en Quebec.

Reference: These tissues are analyzed, processed and frozen before being stored at Héma-Québec, which manages also the only bank of placental blood in Quebec.

Translation	Rank				
These weavings are analyzed, transformed and frozen before being stored in Hema-Quebec, that negotiates also the public only bank of blood of the umbilical cord in Quebec.	○ 1 Best	○ 2	○ 3	○ 4	● 5 Worst
These tissues analysed, processed and before frozen of stored in Hema-Québec, which also operates the only public bank umbilical cord blood in Quebec.	○ 1 Best	○ 2	● 3	○ 4	○ 5 Worst
These tissues are analyzed, processed and frozen before being stored in Hema-Québec, which also manages the only public bank umbilical cord blood in Quebec.	○ 1 Best	● 2	○ 3	○ 4	○ 5 Worst
These tissues are analyzed, processed and frozen before being stored in Hema-Quebec, which also operates the only public bank of umbilical cord blood in Quebec.	● 1 Best	○ 2	○ 3	○ 4	○ 5 Worst
These fabrics are analyzed, are transformed and are frozen before being stored in Hema-Québec, who manages also the only public bank of blood of the umbilical cord in Quebec.	○ 1 Best	○ 2	○ 3	● 4	○ 5 Worst

Correlation with Human Judgments

- Kendall's Tau

$$\tau = \frac{\text{num concordant pairs - num discordant pairs}}{\text{total pairs}}$$

You implemented your own evaluation metric

Noisy Channel Model

$$
\begin{aligned}
\mathbf{e}^* &= \arg\max_{\mathbf{e}} p(\mathbf{e} \mid \mathbf{g}) \\
&= \arg\max_{\mathbf{e}} \frac{p(\mathbf{g} \mid \mathbf{e}) \times p(\mathbf{e})}{p(\mathbf{g})} \\
&= \arg\max_{\mathbf{e}} p(\mathbf{g} \mid \mathbf{e}) \times p(\mathbf{e}) \\
&= \arg\max_{\mathbf{e}} \log p(\mathbf{g} \mid \mathbf{e}) + \log p(\mathbf{e})
\end{aligned}
$$

Noisy Channel Model

$$
\begin{aligned}
\mathbf{e}^* &= \arg\max_{\mathbf{e}} p(\mathbf{e} \mid \mathbf{g}) \\
&= \arg\max_{\mathbf{e}} \frac{p(\mathbf{g} \mid \mathbf{e}) \times p(\mathbf{e})}{p(\mathbf{g})} \\
&= \arg\max_{\mathbf{e}} p(\mathbf{g} \mid \mathbf{e}) \times p(\mathbf{e}) \\
&= \arg\max_{\mathbf{e}} \log p(\mathbf{g} \mid \mathbf{e}) + \log p(\mathbf{e}) \\
&= \boxed{\arg\max_{\mathbf{e}} \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}^\top}_{\mathbf{w}^\top} \underbrace{\begin{bmatrix} \log p(\mathbf{g} \mid \mathbf{e}) \\ \log p(\mathbf{e}) \end{bmatrix}}_{\mathbf{h}(\mathbf{g},\mathbf{e})}}
\end{aligned}
$$

Log-linear Model

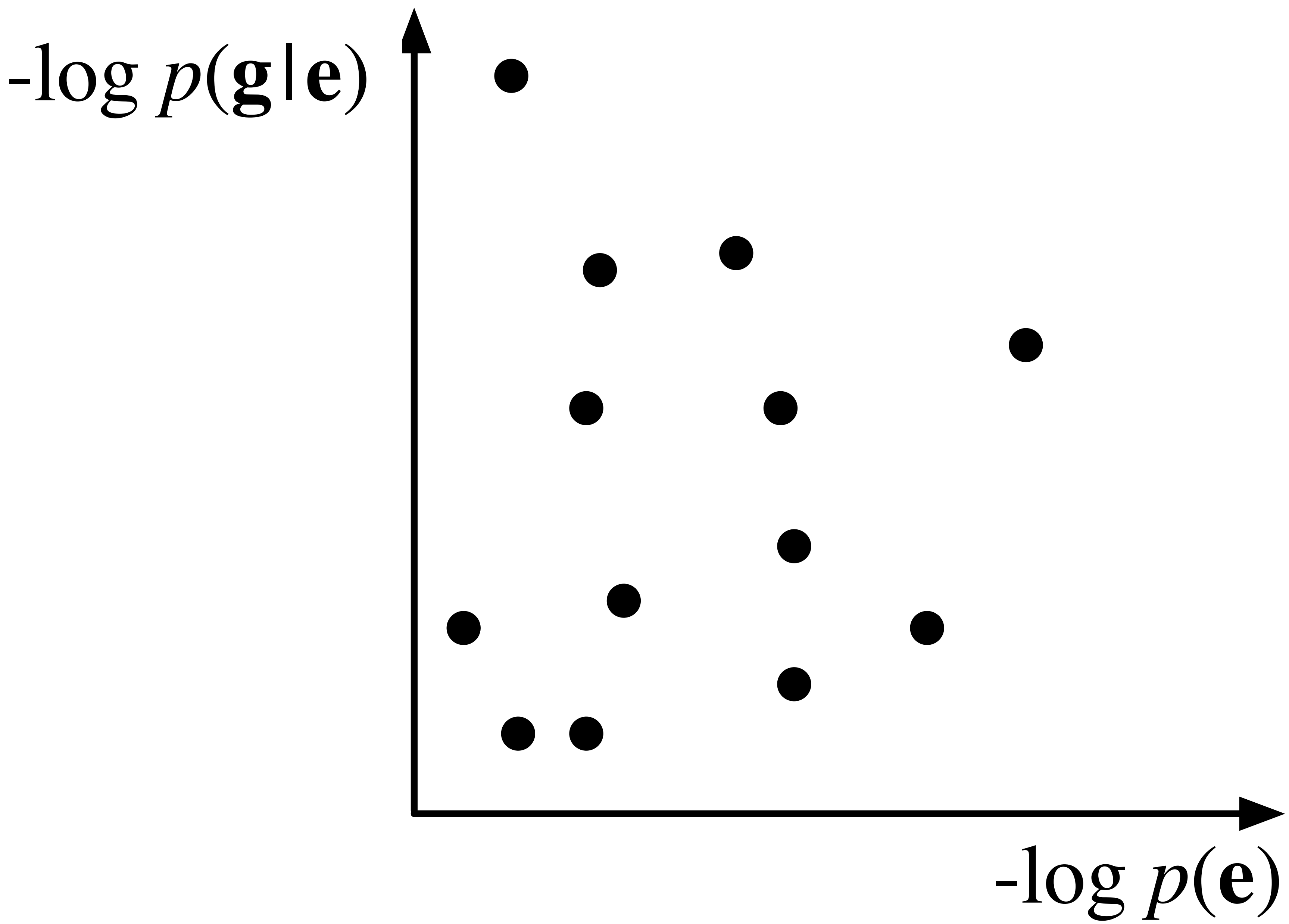
Discriminative modeling

- Depart from generative modeling
- Goal:
    - Directly optimize for translation performance by discriminating between good/bad translation

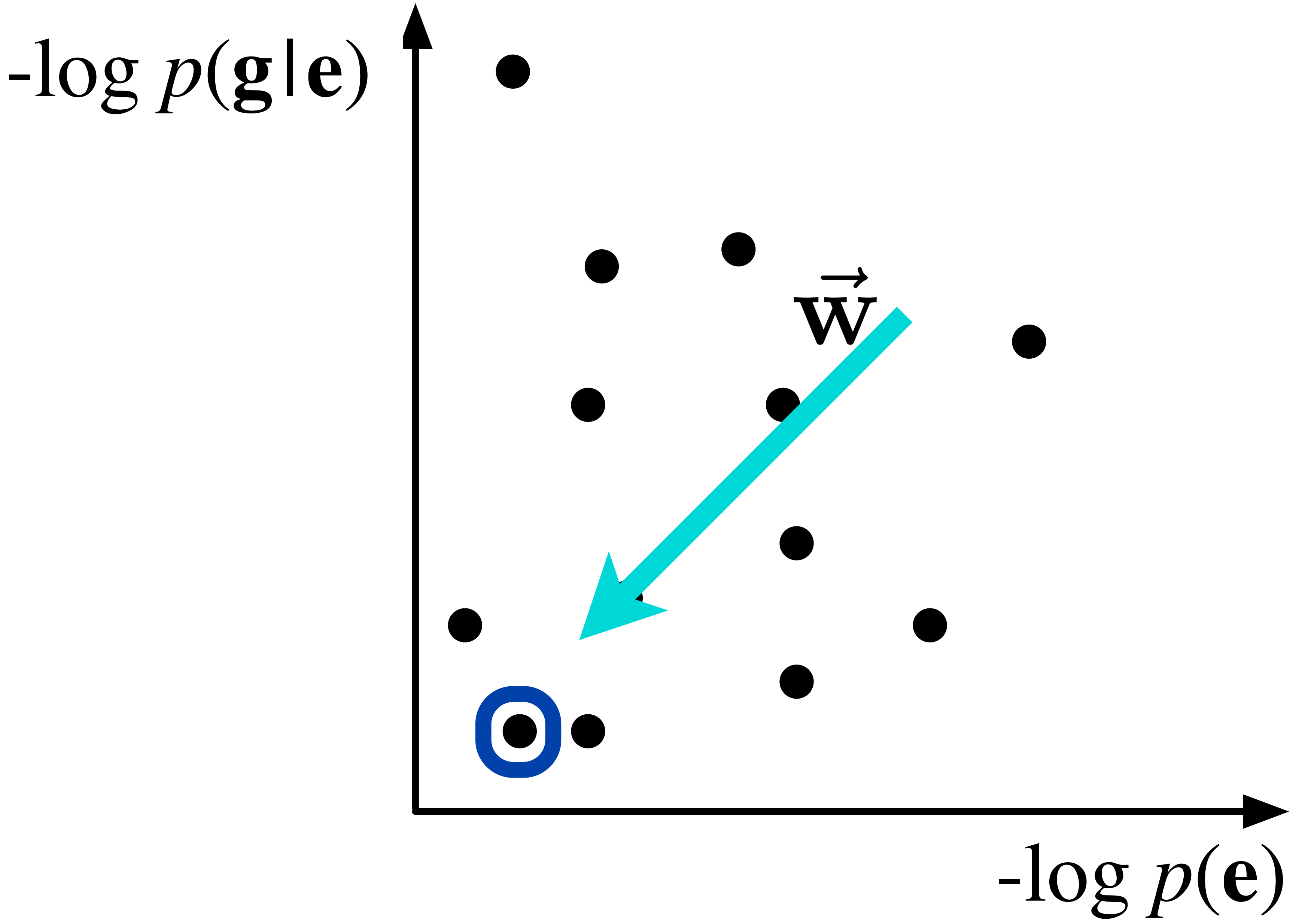
Discriminative modeling

- Represent Possible translations using a set of features **h**
- Each feature $\mathbf{h}_i$ derives from one property of the translation
- Its feature weight $\mathbf{w}_i$ indicates its relative importance

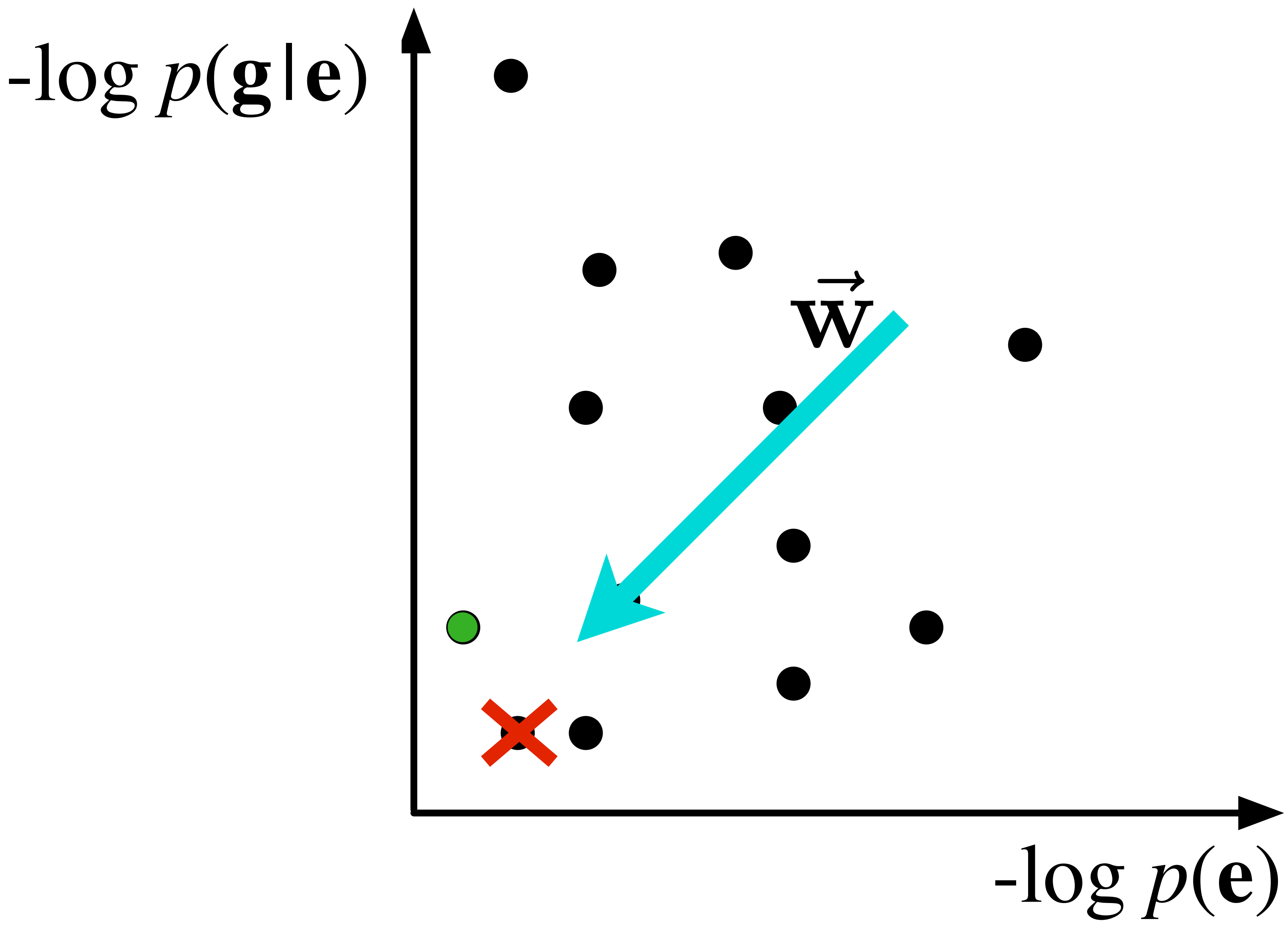
The Noisy Channel



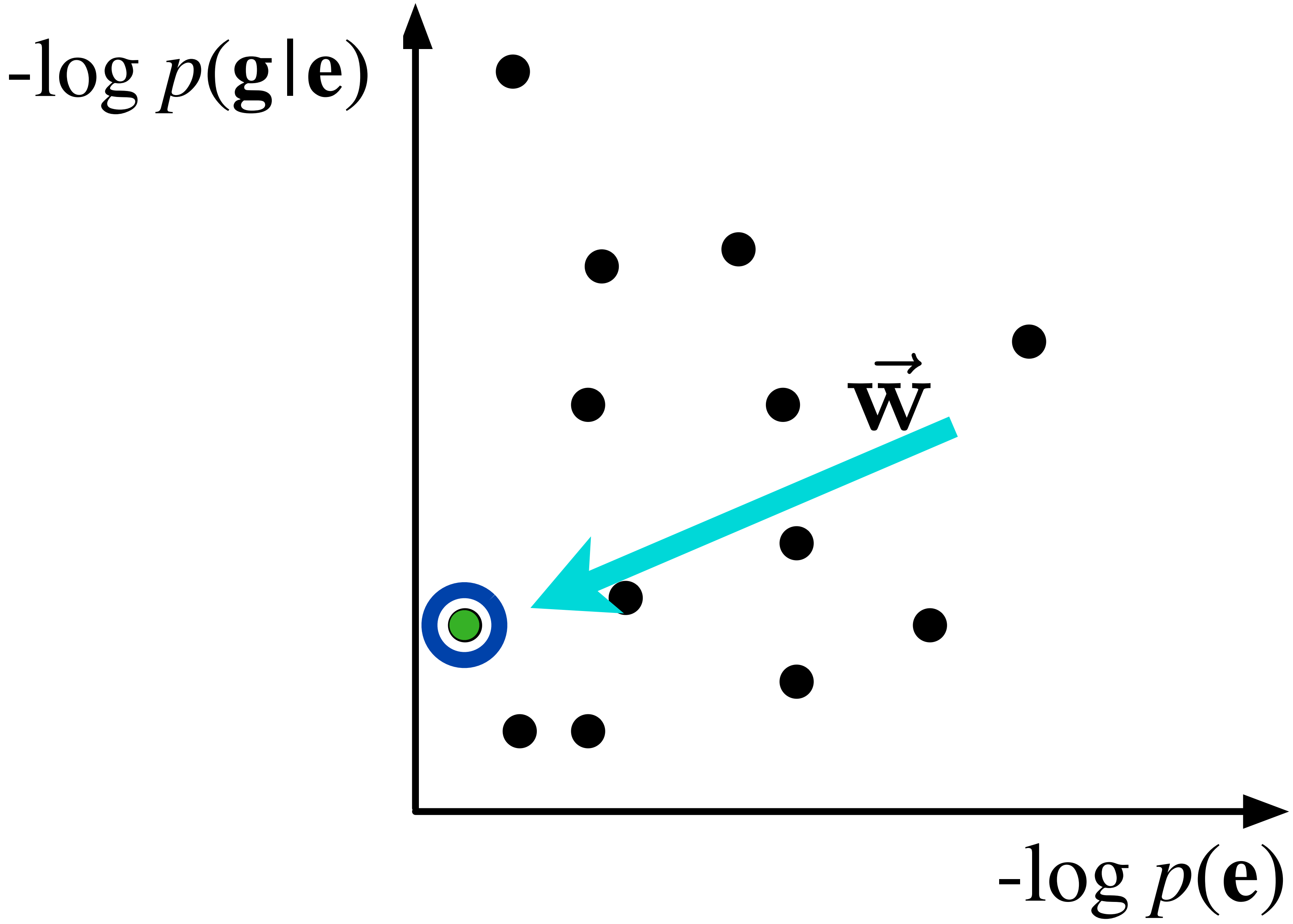
As a Linear Model



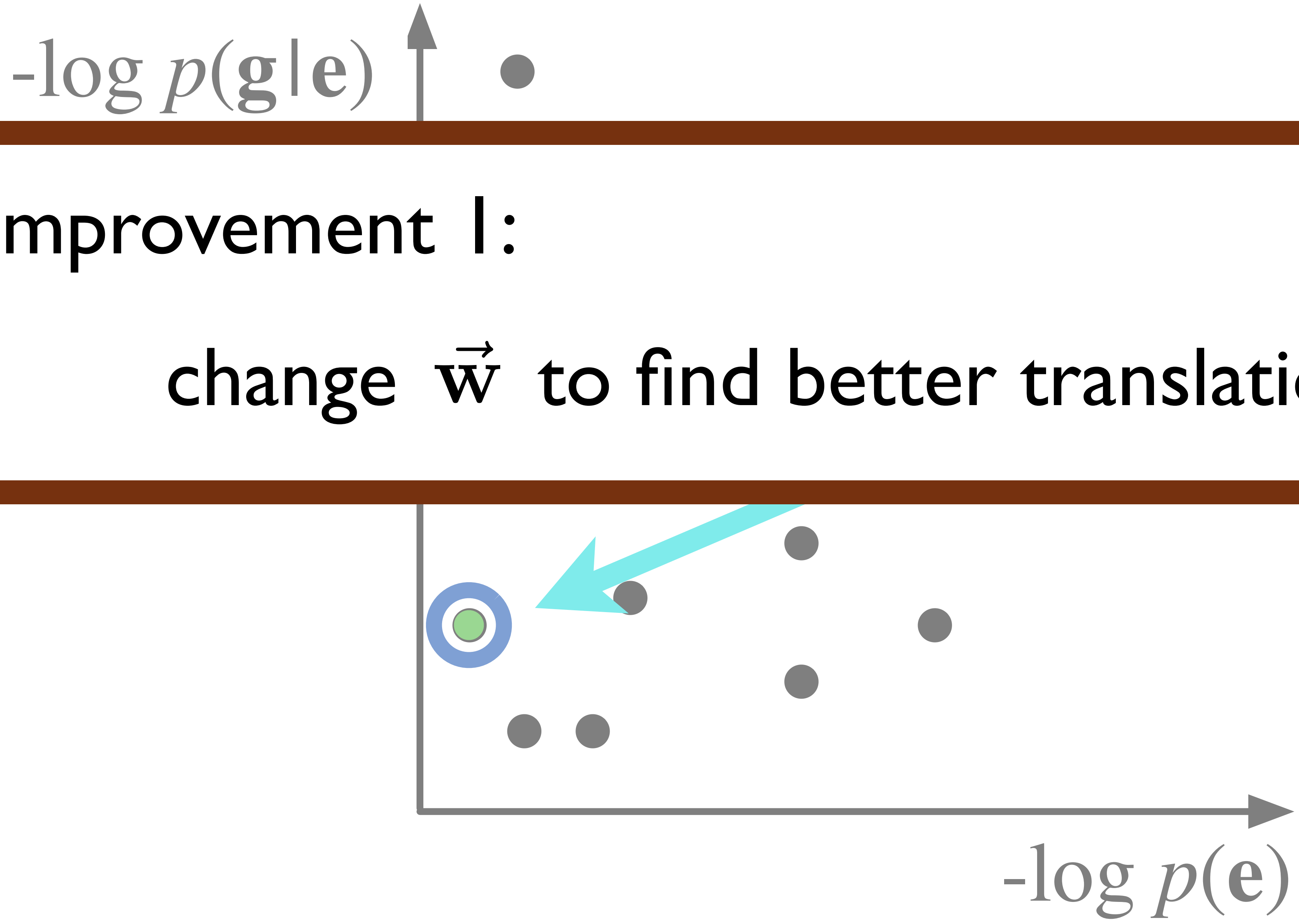
As a Linear Model



As a Linear Model



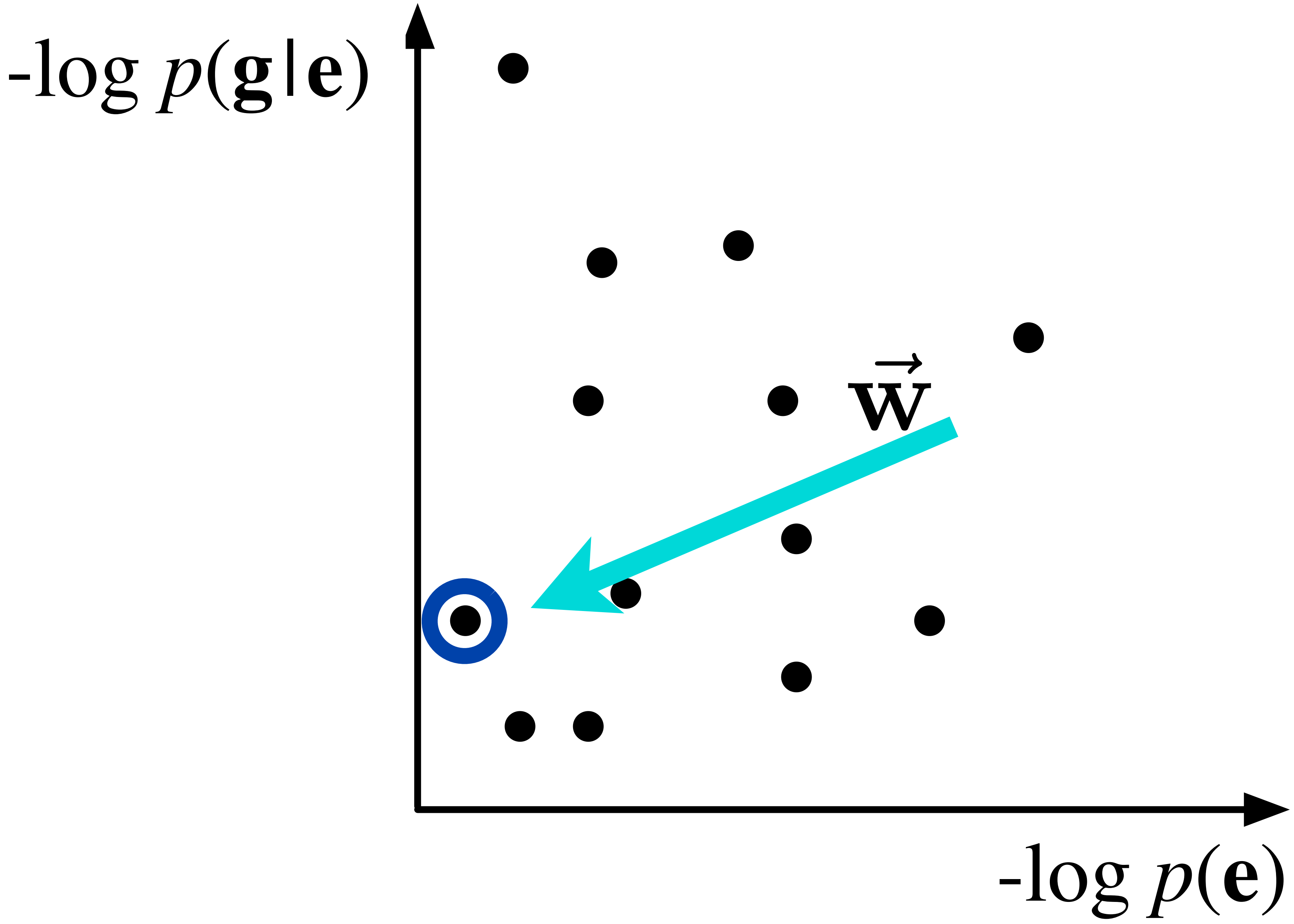
As a Linear Model



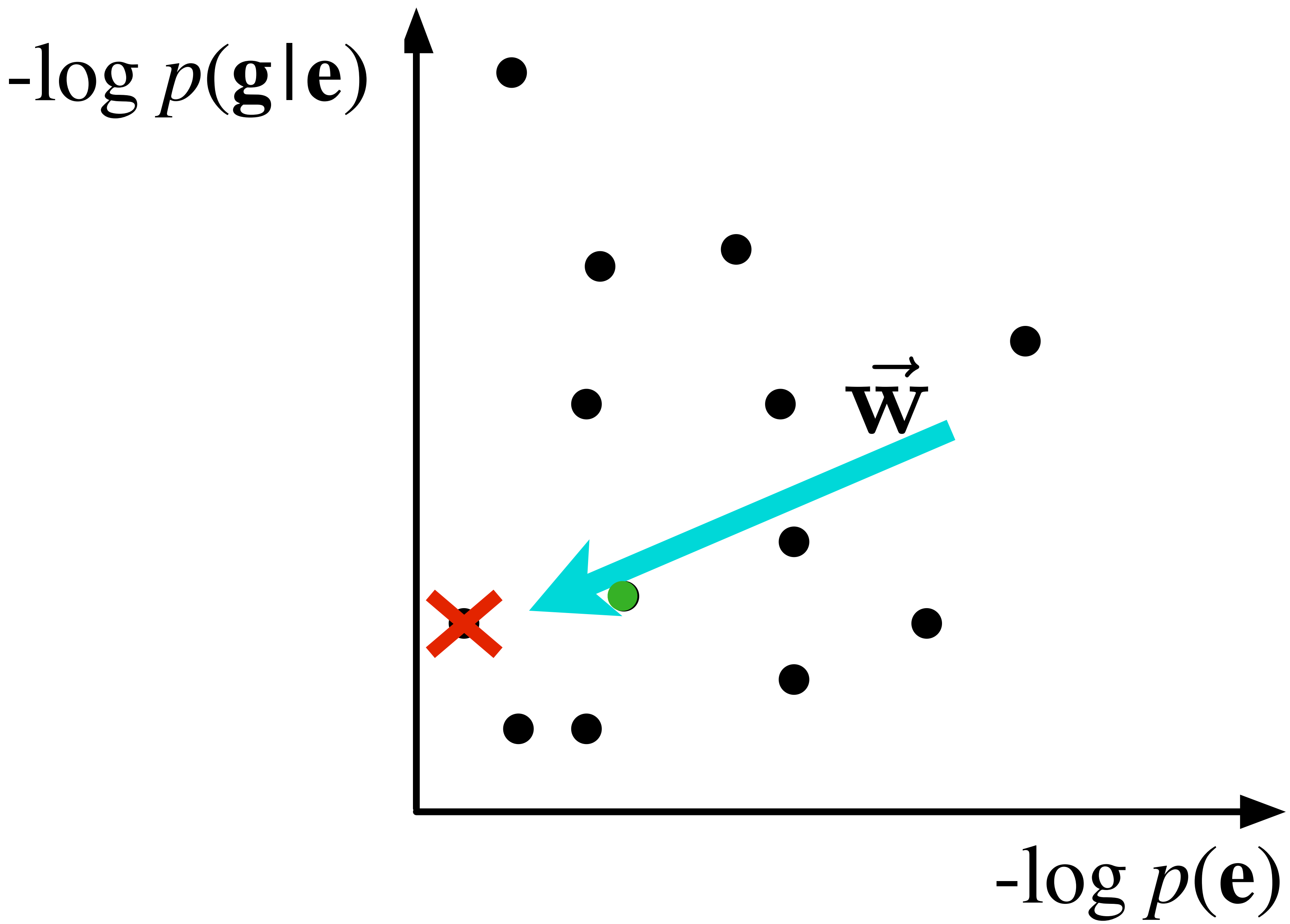
Improvement 1:

change $\vec{w}$ to find better translations

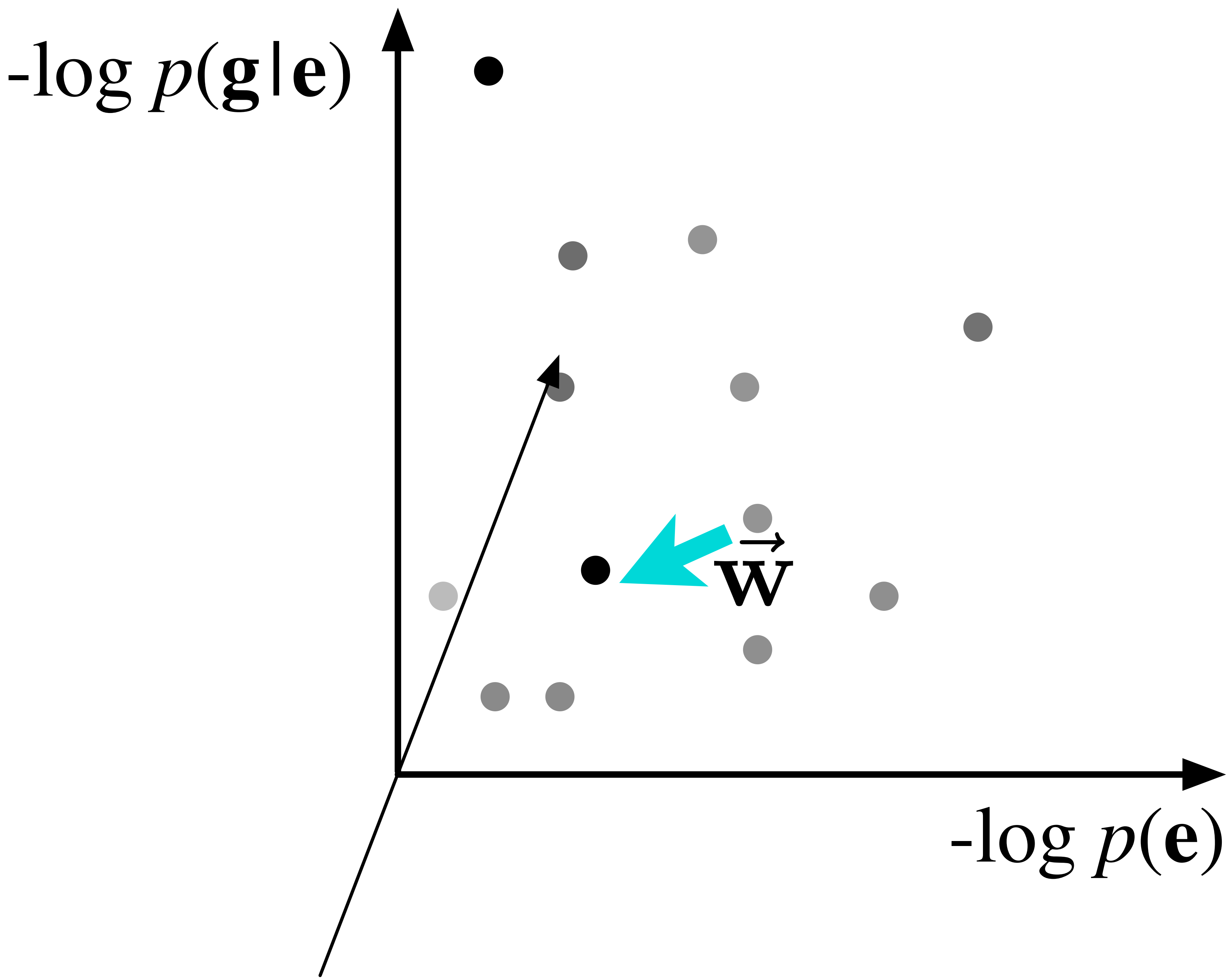
As a Linear Model



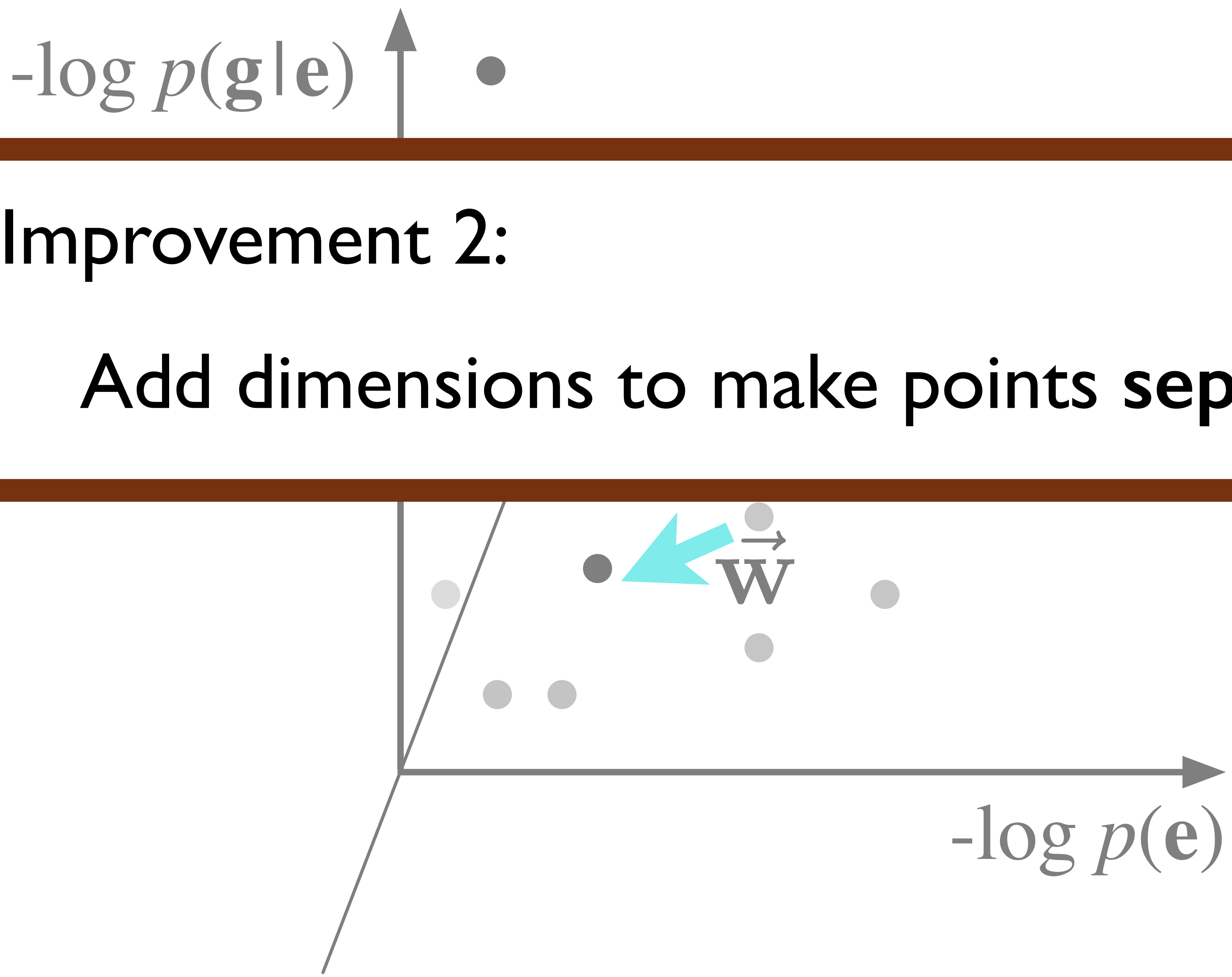
As a Linear Model



As a Linear Model



As a Linear Model

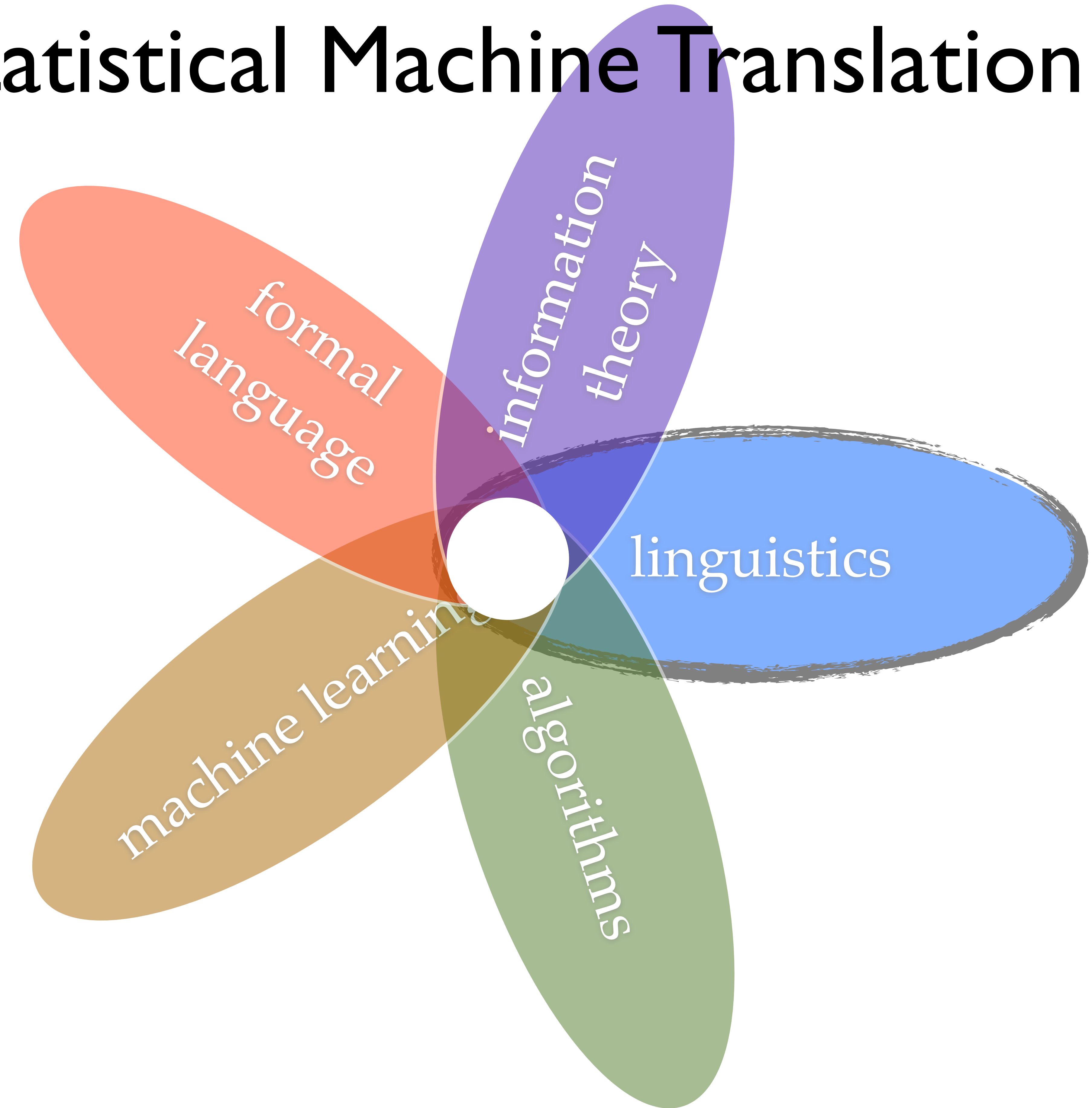


Improvement 2:

Add dimensions to make points **separable**

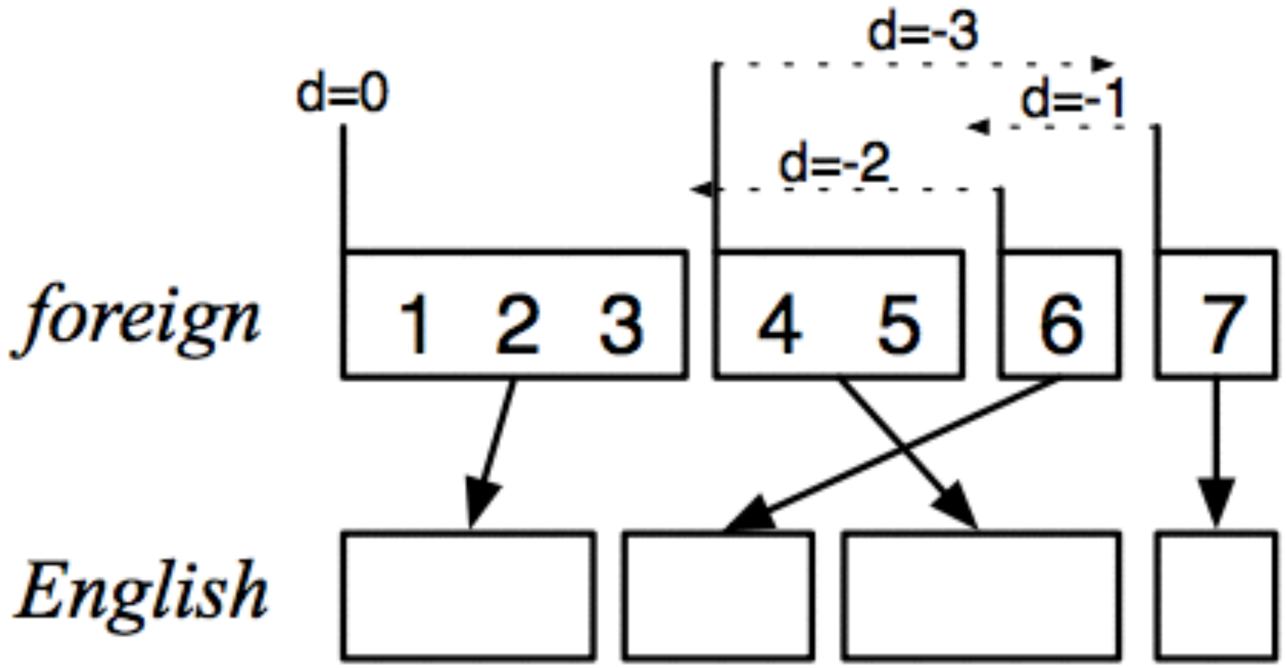
You discriminatively re-ranked n-best translations

Statistical Machine Translation



- Longstanding debate about wheth
  information can help statistical tra
- Two camps

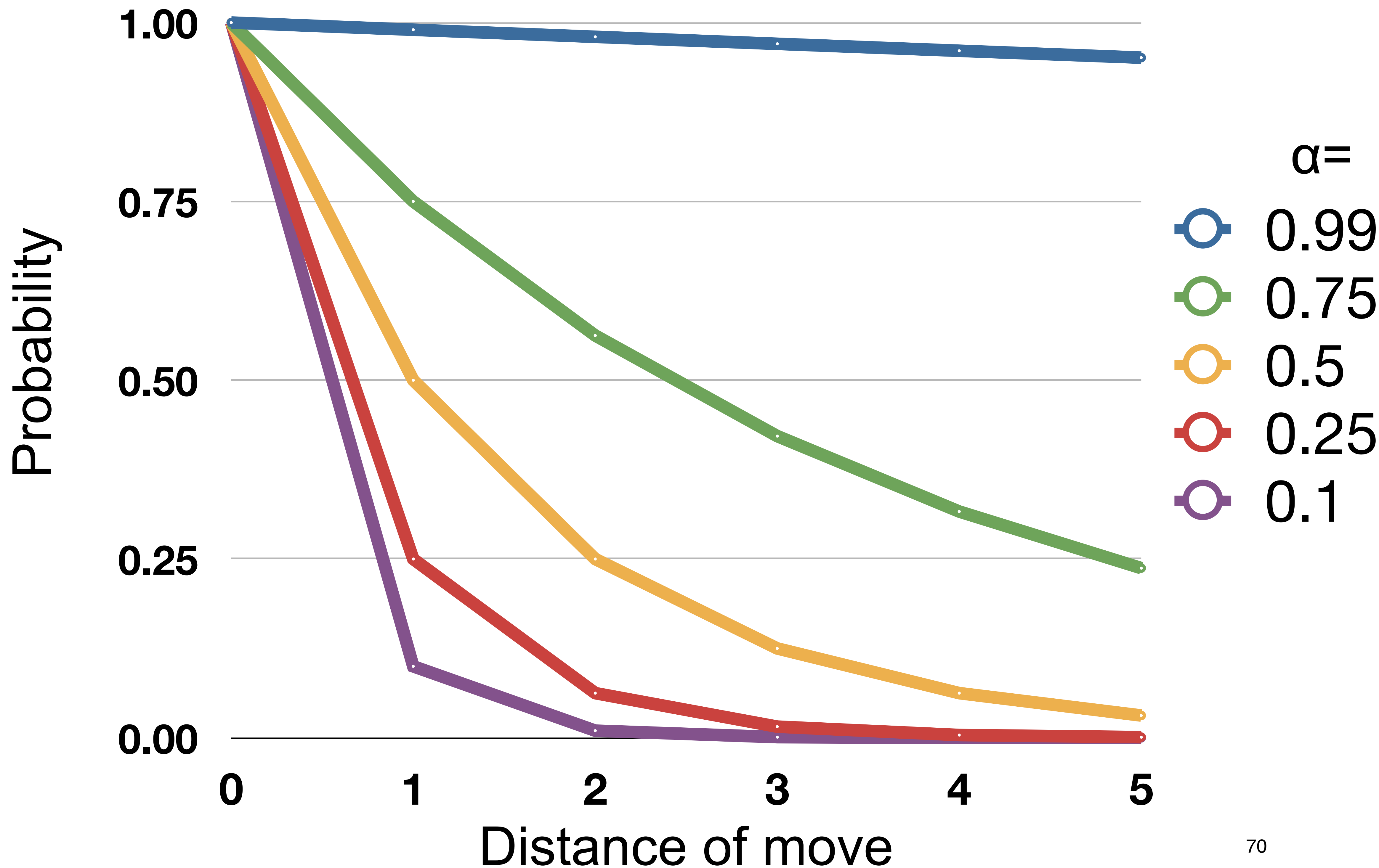
Reordering Model



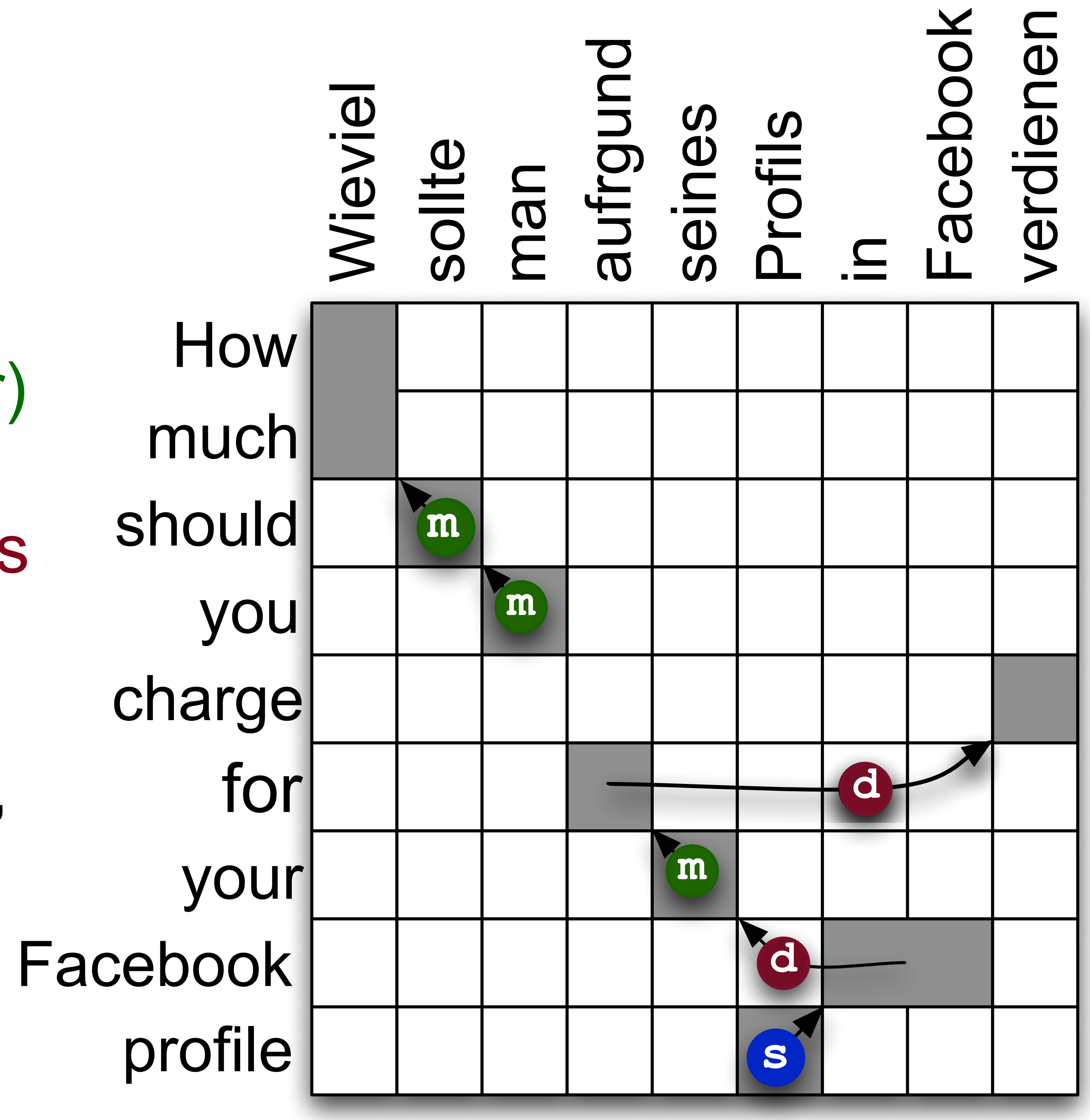
phrase	translates	movement	distance
1	1–3	start at beginning	0
2	6	skip over 4–5	+2
3	4–5	move back over 4–6	-3
4	7	skip over 6	+1

Scoring function: $d(x) = \alpha^{|x|}$ — exponential with distance

Values of α


1.00
0.75
0.50
0.25
0.00
Probability
0
1
2
3
4
5
Distance of move
α=
0.99
0.75
0.5
0.25
0.1

Lexicalized Reordering model



m: monotone (keep order)
s: swap order
d: become discontinuous

Reordering features are probability estimates of s, d, and m

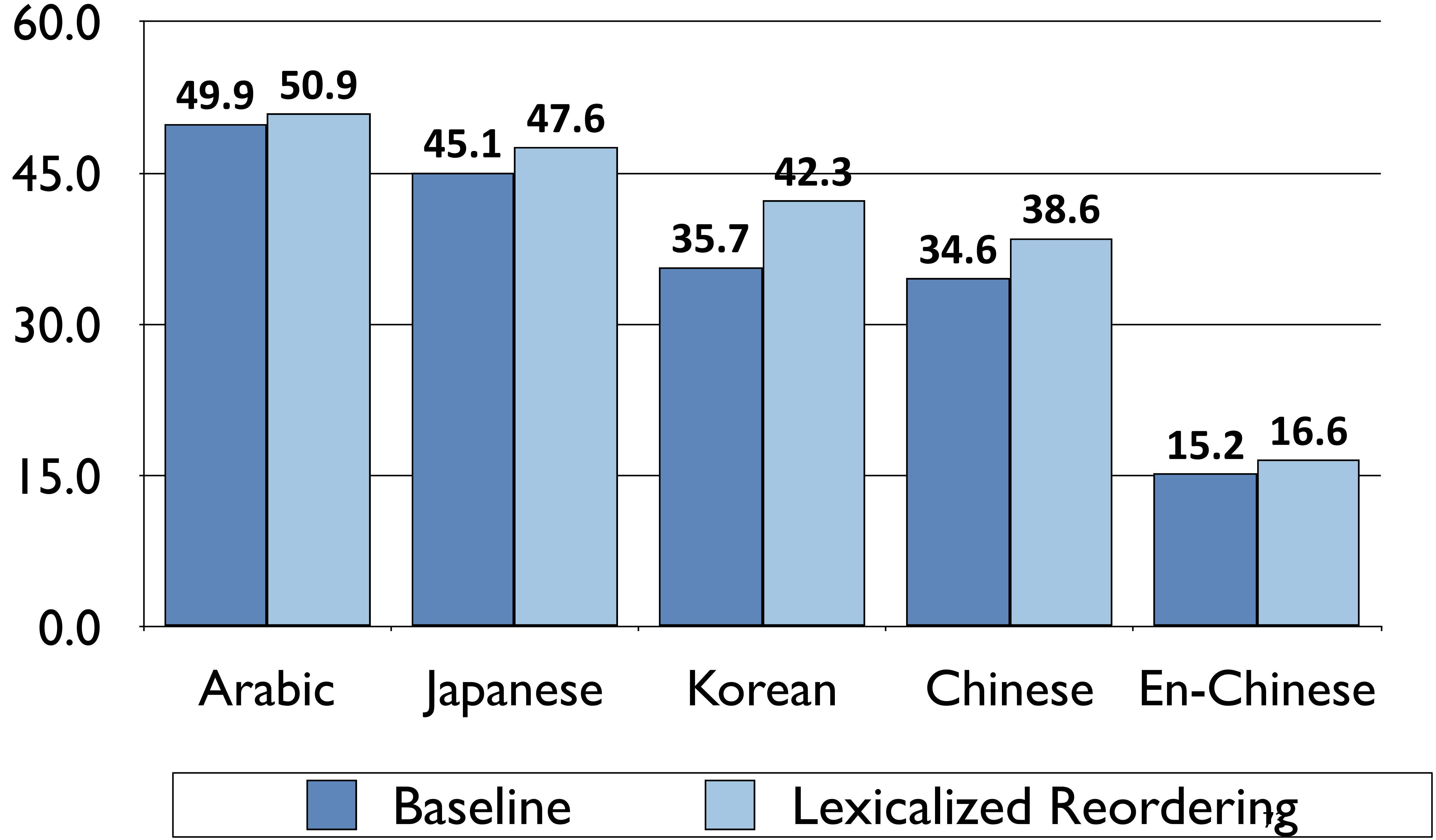
Lexicalized Reordering table

- Identical phrase pairs <f,e> as in the phrase translation table
- Contains values for p(monotone|e,f), p(swap|e,f), p(discontinuous|e,f)

Source	**Translation**	p(m|e,f)	p(s|e,f)	p(d|e,f)
natuerlich	of course	0.52	0.08	0.4
natuerlich	naturally	0.42	0.1	0.48
natuerlich	of course ,	0.5	0.001	0.499
natuerlich	, of course	0.27	0.17	0.56

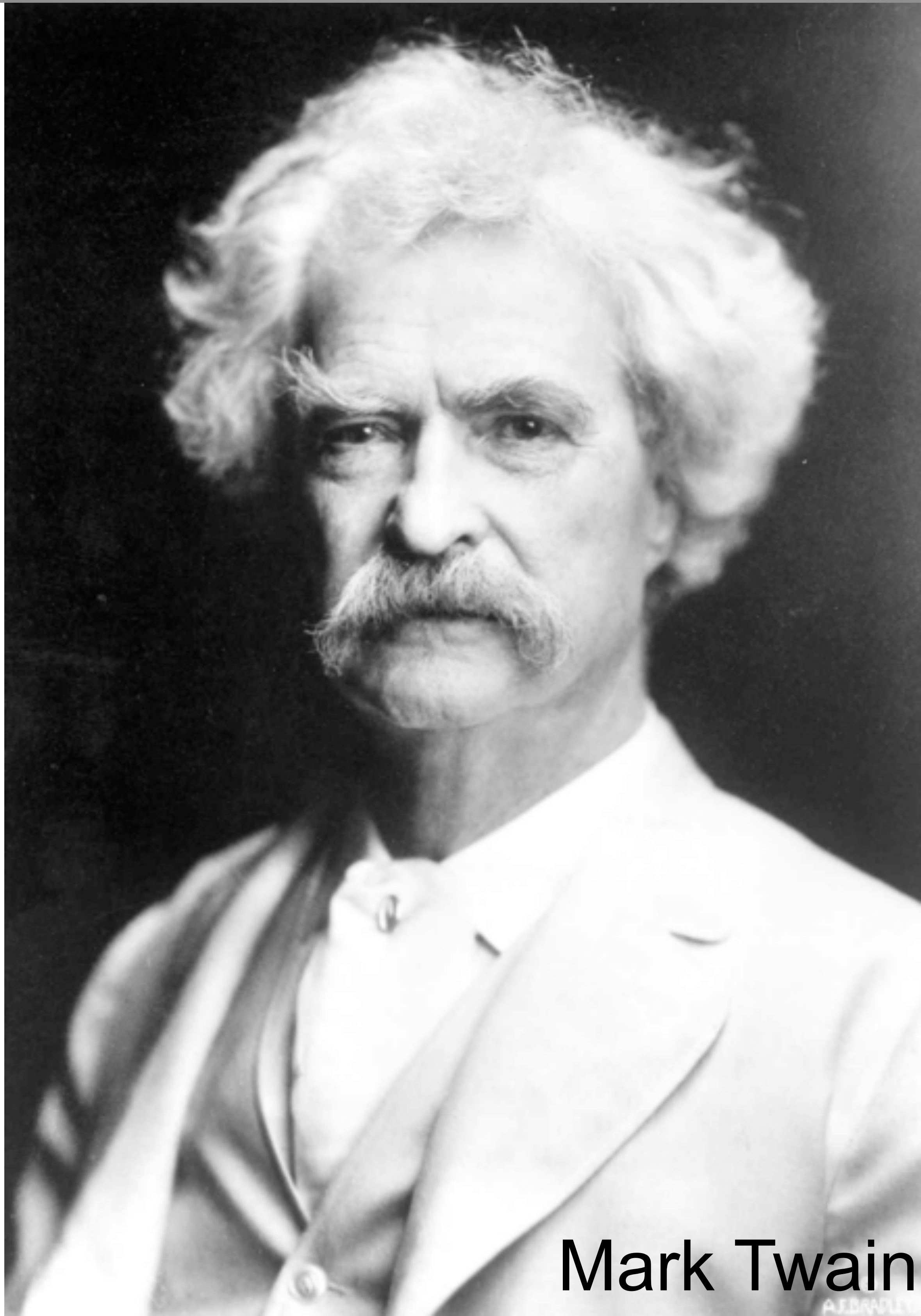
Empirically Better

Koehn et al, IWSLT 2005



The Awful German Language

"The Germans have another kind of parenthesis, which they make by splitting a verb in two and putting half of it at the beginning of an exciting chapter and the OTHER HALF at the end of it. Can any one conceive of anything more confusing than that? These things are called 'separable verbs.' The wider the two portions of one of them are spread apart, the better the author of the crime is pleased with his performance."



German verbs

Ich werde Ihnen den Report aushaendigen .
I will to_you the report pass_on .

Ich werde Ihnen die entsprechenden Anmerkungen aushaendigen .
I will to_you the corresponding comments pass_on .

Ich werde Ihnen die entsprechenden Anmerkungen am Dienstag aushaendigen
I will to_you the corresponding comments on Tuesday pass_on

Collins' Pre-ordering Model

Step 1: Reorder the source language

Ich werde Ihnen den Report aushaendigen ,
damit Sie den eventuell uebernehmen koennen .

Ich werde aushaendigen Ihnen den Report ,
damit Sie koennen uebernehmen den eventuell .

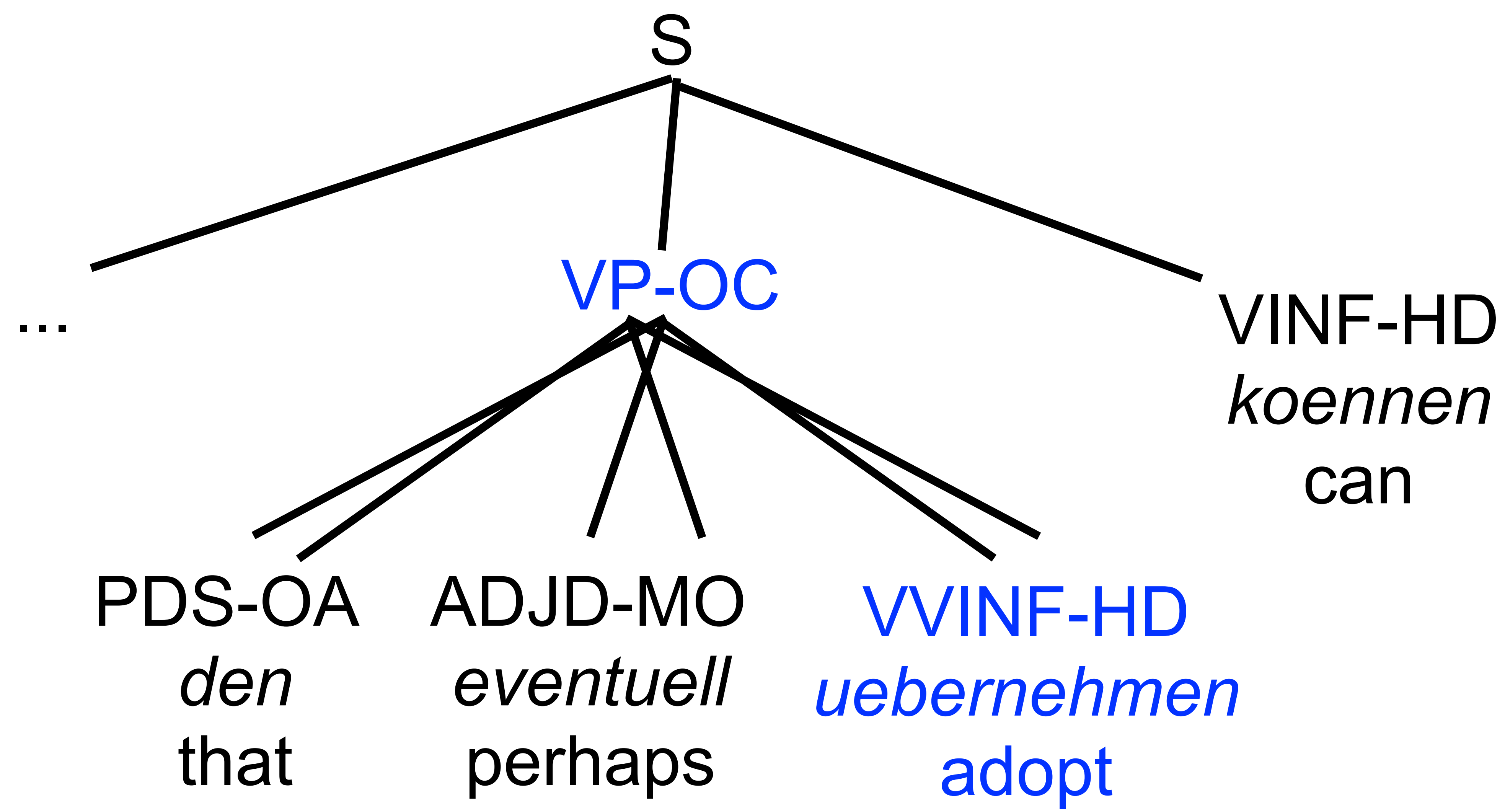
(I will pass_on to_you the report, so_that you can adopt it perhaps .)

Step 2: Apply the phrase-based machine translation pipeline to the reordered input.

Clause Restructuring

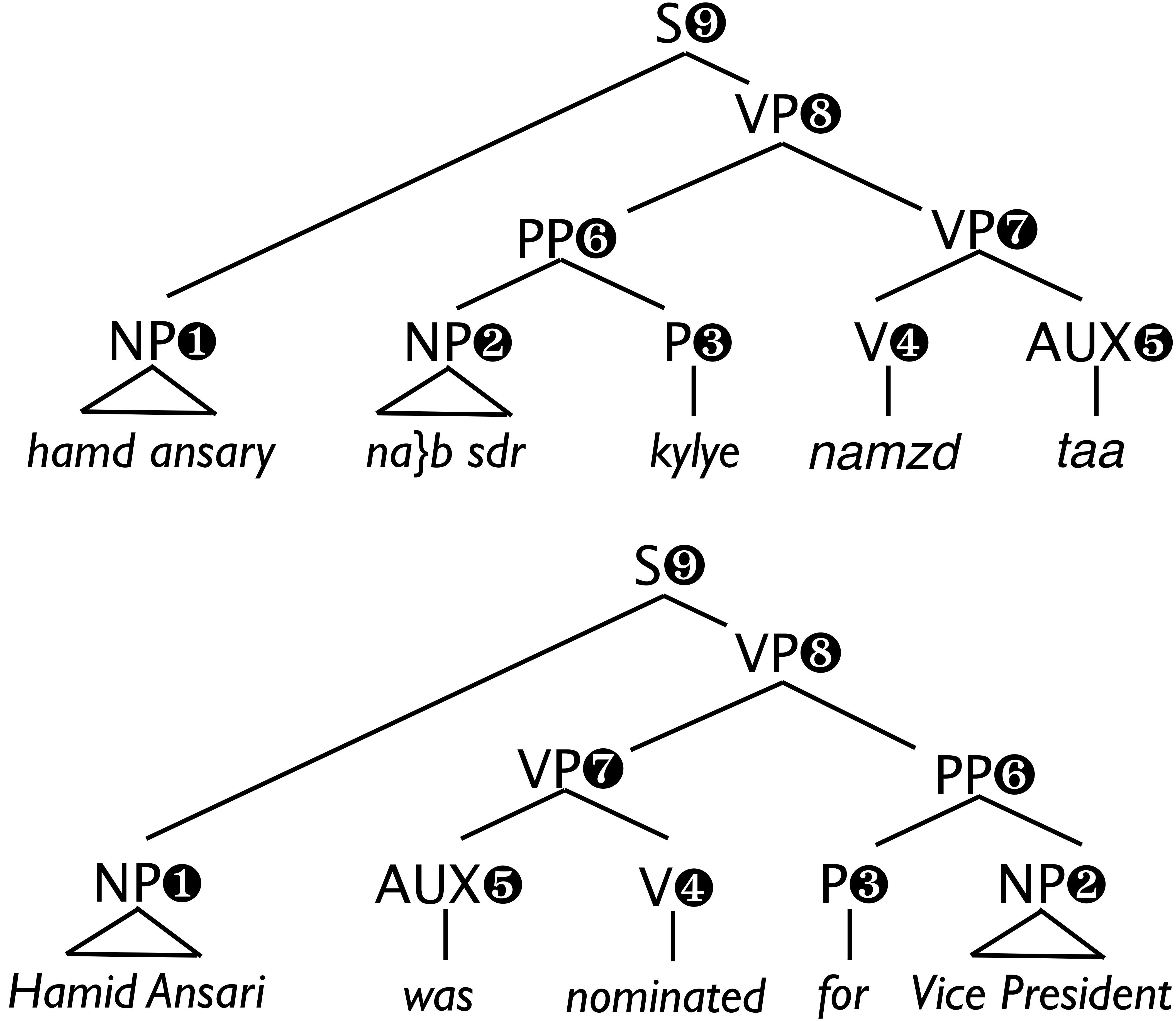
Rule 1: **Verbs are initial in VPs**

Within a VP, move the head to the initial position



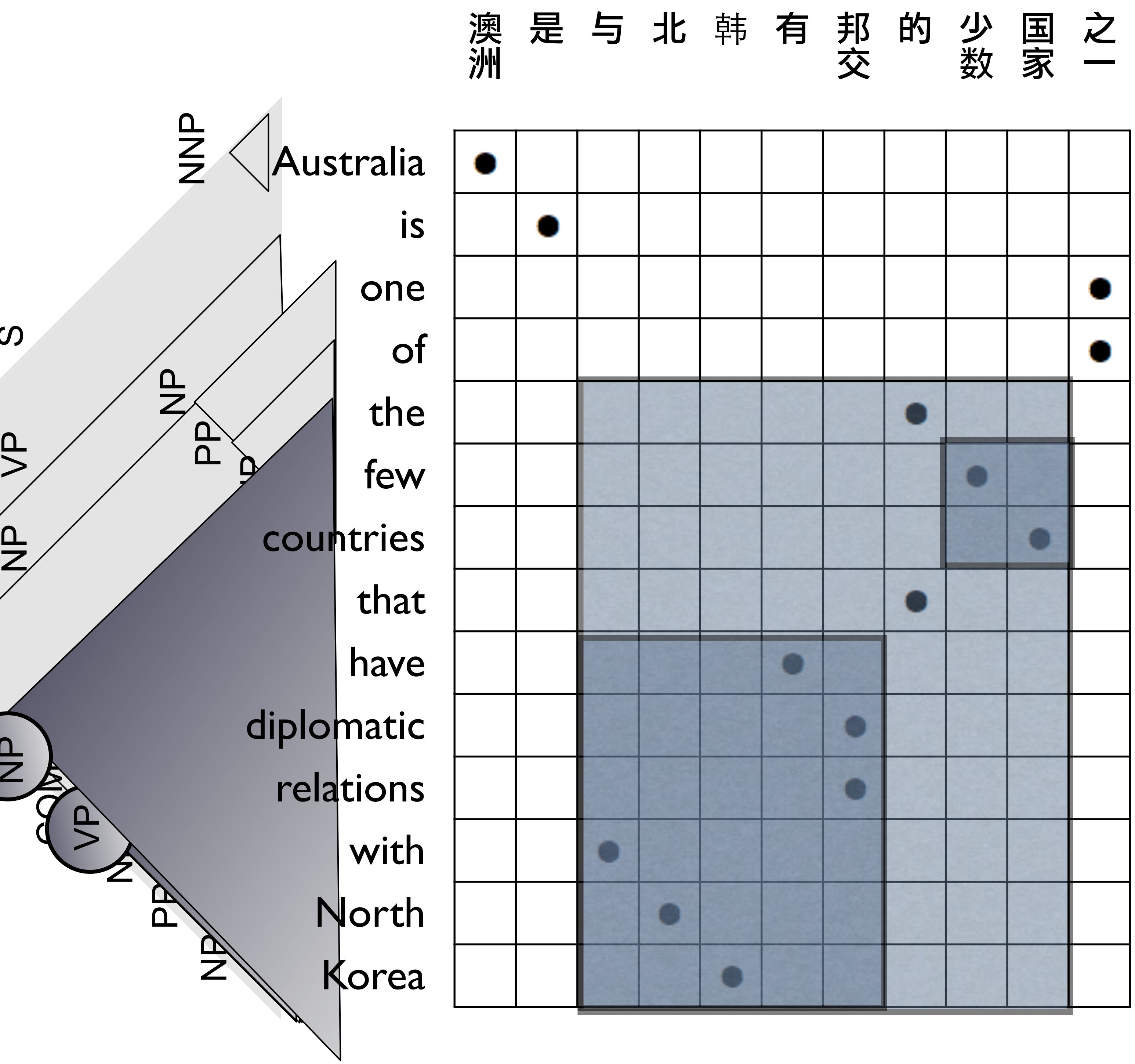
Synchronous Context Free Grammars

- A common way of representing syntax in NLP is through context free grammars
- Synchronous context free grammars generate pairs of corresponding strings
- Can be used to describe translation and re-ordering between languages
- SCFGs translate sentences by parsing them



S❾
VP❽
PP❻
VP❼
NP❶
NP❷
P❸
V❹
AUX❺
hamd ansary
na}b sdr
kylye
namzd
taa
S❾
VP❽
VP❼
PP❻
NP❶
AUX❺
V❹
P❸
NP❷
Hamid Ansari
was
nominated
for
Vice President

Extracting Syntactic Rules



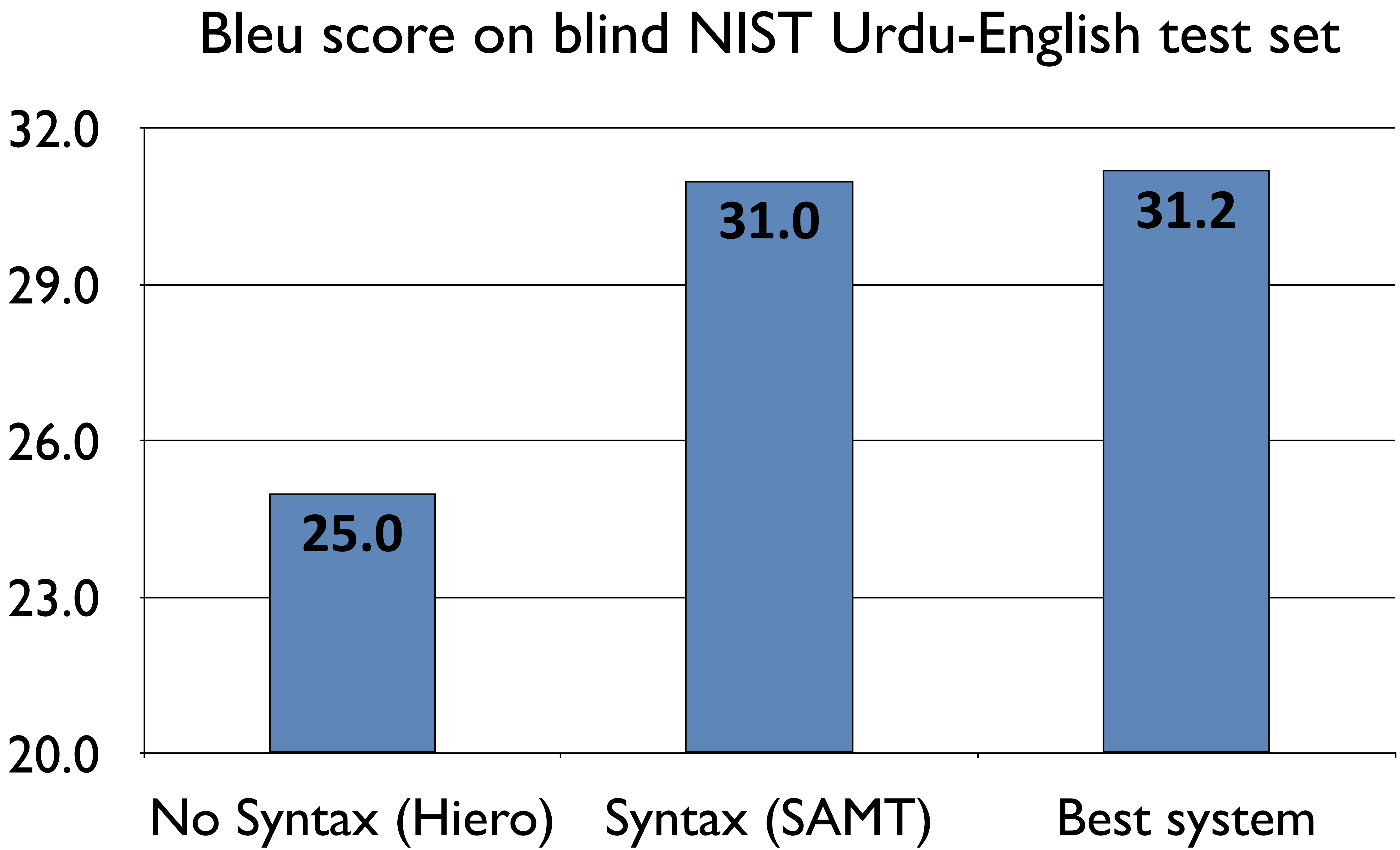
VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

NP → 与 北 韩 有 邦交 的 少数 国家, the few countries that have diplomatic relations with North Korea

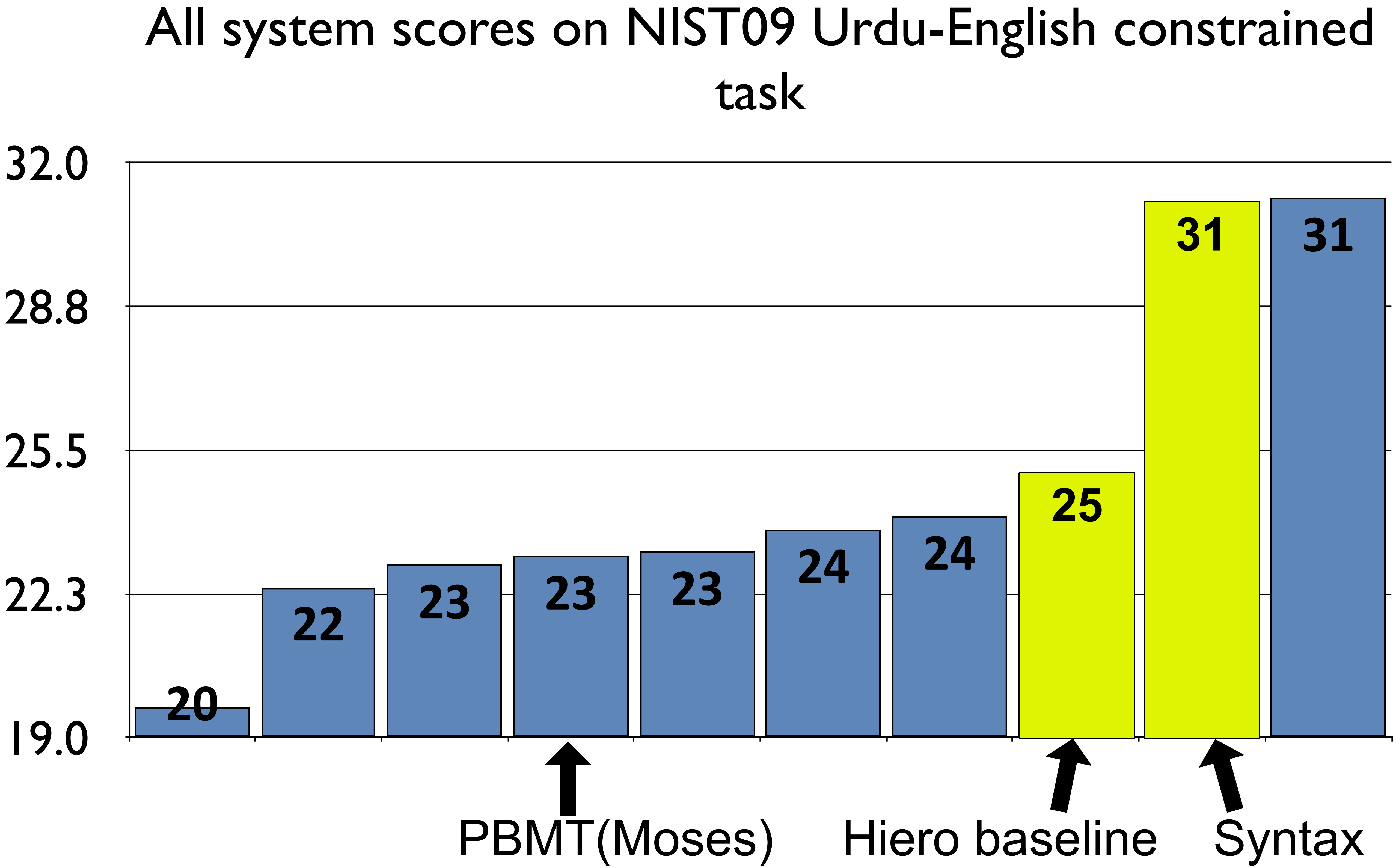
NP → VP 的 少数 国家, the few countries that VP

NP → VP 的 NP, the NP that VP

Syntax v. no Syntax



State of the Art Urdu Results



Joshua Decoder

- An open source decoder
- Uses synchronous context free grammars to translate
- Implements all algorithms needed for translating with SCFGs
  - grammar extraction (Thrax!)
  - chart-parsing
  - n-gram language model integration
  - pruning, and k-best extraction

History of Decoders

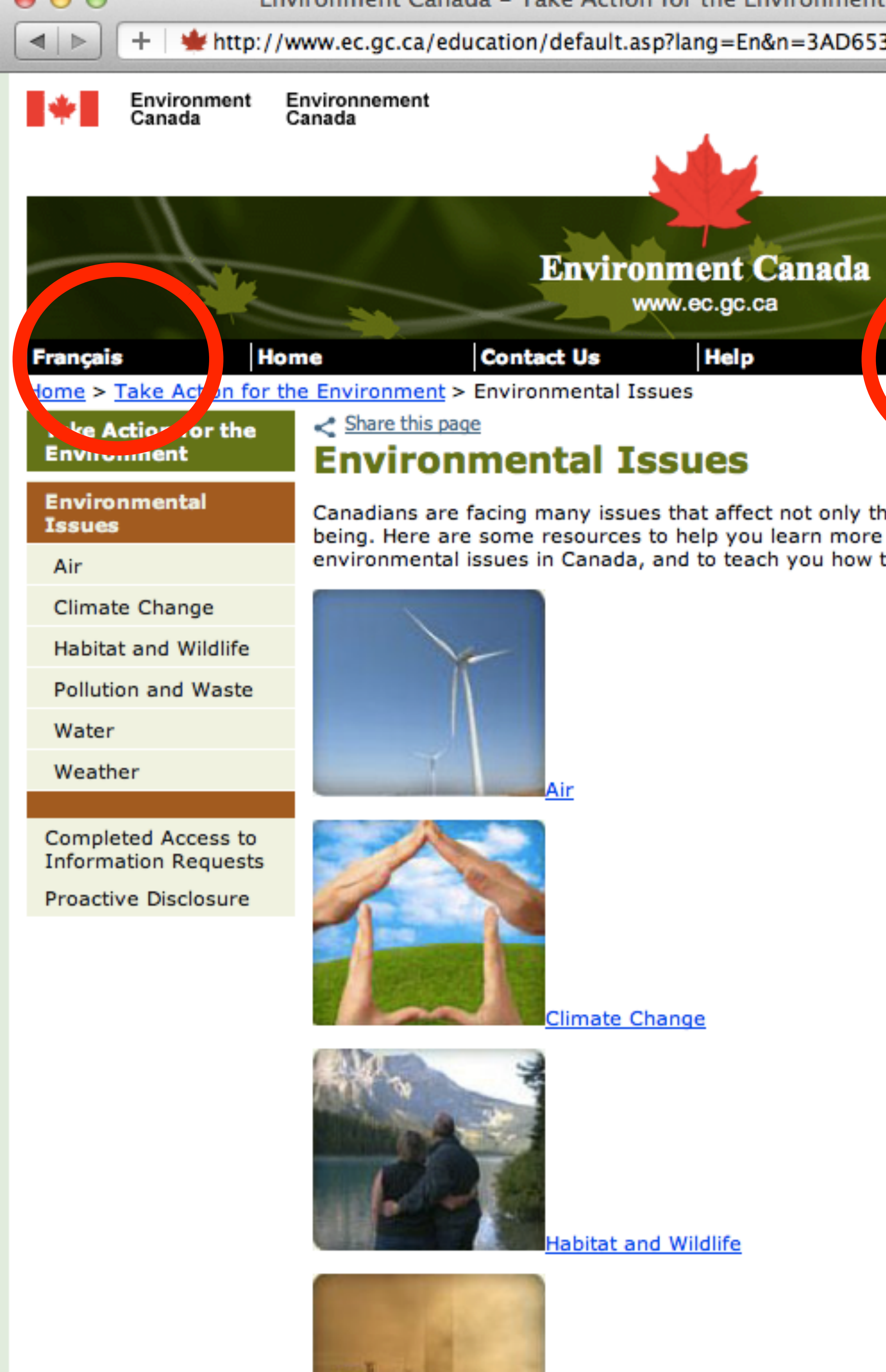
GIZA++ was an open source implementation of the IBM alignment models developed at the 1999 CLSP summer workshop

PHARAOH was a beam search decoder for phrase-based statistical machine translation models

Moses is an open source decoder for phrase-based statistical machine translation models

JosHUa is an open source syntax based SMT models

Advanced Topics


Environment Canada – Take Action for the Environment
http://www.ec.gc.ca/education/default.asp?lang=En&n=3AD653
Environment Canada
Environnement Canada
Environment Canada
www.ec.gc.ca
Français
Home
Contact Us
Help
Home > Take Action for the Environment > Environmental Issues
Take Action for the Environment
Share this page
Environmental Issues
Environmental Issues
Canadians are facing many issues that affect not only th
being. Here are some resources to help you learn more
environmental issues in Canada, and to teach you how t
Air
Climate Change
Habitat and Wildlife
Pollution and Waste
Water
Weather
Completed Access to Information Requests
Proactive Disclosure
Air
Climate Change
Habitat and Wildlife


Environnement Canada – Passons à l'action pour l'environnement
http://www.ec.gc.ca/education/default.asp?lang=Fr&n=3AD65317-1
Environment Canada
Environnement Canada
Environnement Canada
www.ec.gc.ca
English
Accueil
Contactez-nous
Aide
Accueil > Passons à l'action pour l'environnement > Questions sur l'environnement
Passons à l'action pour l'environnement
Partagez cette page
Questions sur l'environnement
Questions sur l'environnement
Les Canadiens font face à plusieurs questions concernant no
leur santé et leur bien-être. Voici quelques moyens qui vous
causes et effets des grands enjeux environnementaux au Ca
prendre ces mesures et les appliquer.
L'air
Changement climatique
Habitat et faune
Pollution et déchets
L'eau
La météo
Demandes d'accès à l'information complétées
Divulgation proactive
L'air
Changement climatique
Habitat et faune

Google's efforts

- Number of words of mined English-X parallel text

	baseline	books	web
Czech	27.5M	-	271.9M
French	479.8M	228.5M	4,914.3M
German	54.2M	-	3,787.6M
Hungarian	26.9M	-	198.9M
Spanish	441.0M	15.0M	4,846.8M

- Translation improvements

	baseline	+books	+web
Czech English	21.59	-	29.26 (+7.67)
German English	27.99	-	32.35 (+4.36)
French English	34.26	34.73 (+0.47)	36.65 (+2.39)
Spanish English	43.67	44.07 (+0.40)	46.21 (+2.54)

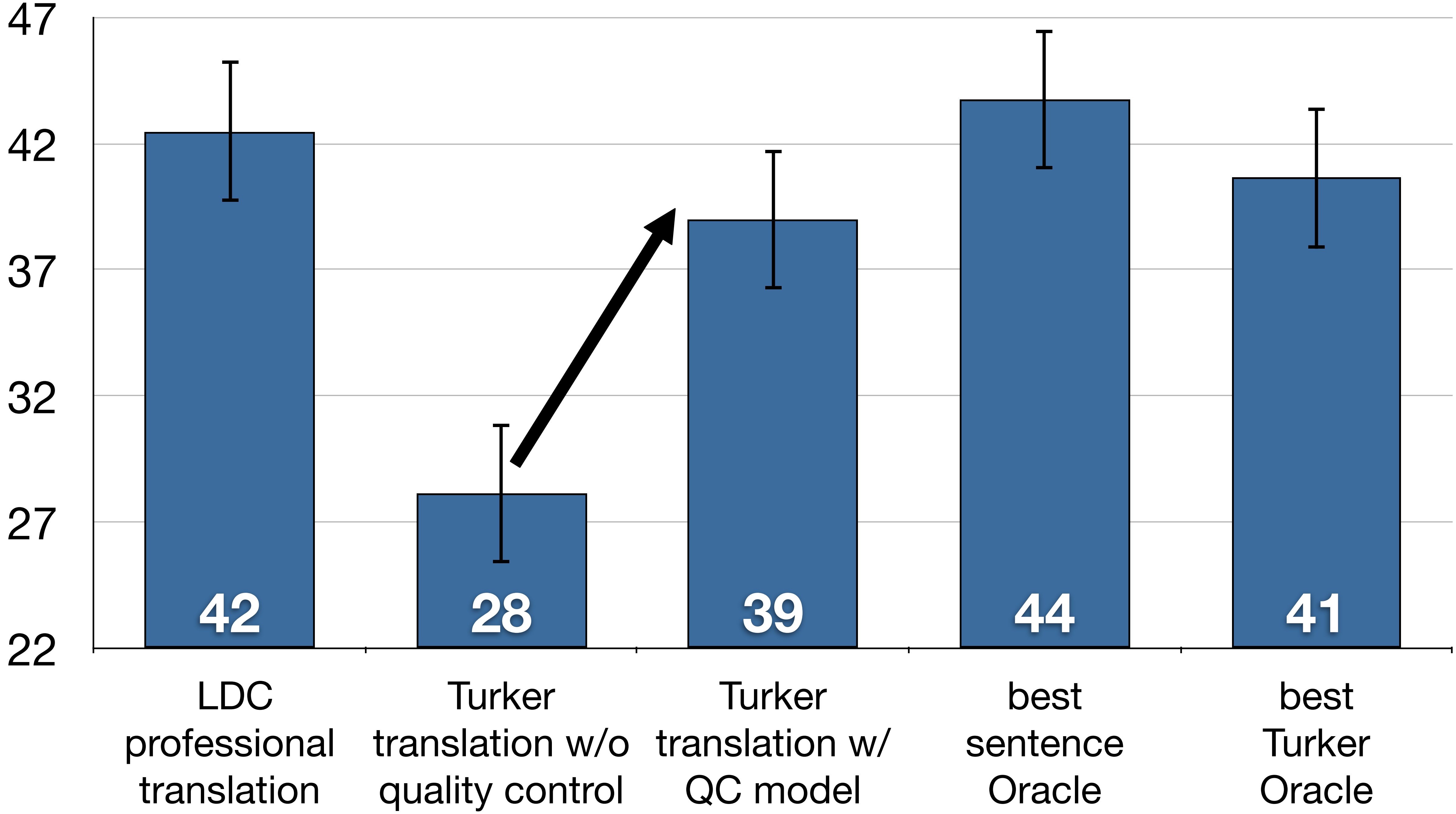
Avoiding dieting to prevent from flu	abstention from dieting in order to avoid Flu	Abstain from decrease eating in order to escape from flue	In order to be safer from flu quit dieting
This research of American scientists came in front after experimenting on mice.	This research from the American Scientists have come up after the experiments on rats.	This research of American scientists was shown after many experiments on mouses.	According to the American Scientist this research has come out after much experimentations on rats.
Experiments proved that mice on a lower calorie diet had comparatively less ability to fight the flu virus.	in has been proven from experiments that rats put on diet with less calories had less ability to resist the Flu virus.	It was proved by experiments the low calories eaters mouses had low defending power for flue in ratio.	Experimentaions have proved that those rats on less calories diet have developed a tendency of not overcoming the flu virus.
research has proven this old myth wrong that its better to fast during fever.	Research disproved the old axiom that " It is better to fast during fever"	The research proved this old talk that decrease eating is useful in fever.	This Research has proved the very old saying wrong that it is good to starve while in fever.

Avoiding dieting to prevent from flu	abstention from dieting in order to avoid Flu	Abstain from decrease eating in order to escape from flue	In order to be safer from flu quit dieting
This research of American scientists came in front after experimenting on mice.	This research from the American Scientists have come up after the experiments on rats.	This research of American scientists was shown after many experiments on mouses.	According to the American Scientist this research has come out after much experimentations on rats.
Experiments proved that mice on a lower calorie diet had comparatively less ability to fight the flu virus.	in has been proven from experiments that rats put on diet with less calories had less ability to resist the Flu virus.	It was proved by experiments the low calories eaters mouses had low defending power for flue in ratio.	Experimentaions have proved that those rats on less calories diet have developed a tendency of not overcoming the flu virus.
research has proven this old myth wrong that its better to fast during fever.	Research disproved the old axiom that " It is better to fast during fever"	The research proved this old talk that decrease eating is useful in fever.	This Research has proved the very old saying wrong that it is good to starve while in fever.

Avoiding dieting to prevent from flu	abstention from dieting in order to avoid Flu	Abstain from decrease eating in order to escape from flue	In order to be safer from flu quit dieting
This research of American scientists came in front after experimenting on mice.	This research from the American Scientists have come up after the experiments on rats.	This research of American scientists was shown after many experiments on mouses.	According to the American Scientist this research has come out after much experimentations on rats.
Experiments proved that mice on a lower calorie diet had comparatively less ability to fight the flu virus.	in has been proven from experiments that rats put on diet with less calories had less ability to resist the Flu virus.	It was proved by experiments the low calories eaters mouses had low defending power for flue in ratio.	Experimentaions have proved that those rats on less calories diet have developed a tendency of not overcoming the flu virus.
research has proven this old myth wrong that its better to fast during fever.	Research disproved the old axiom that " It is better to fast during fever"	The research proved this old talk that decrease eating is useful in fever.	This Research has proved the very old saying wrong that it is good to starve while in fever.

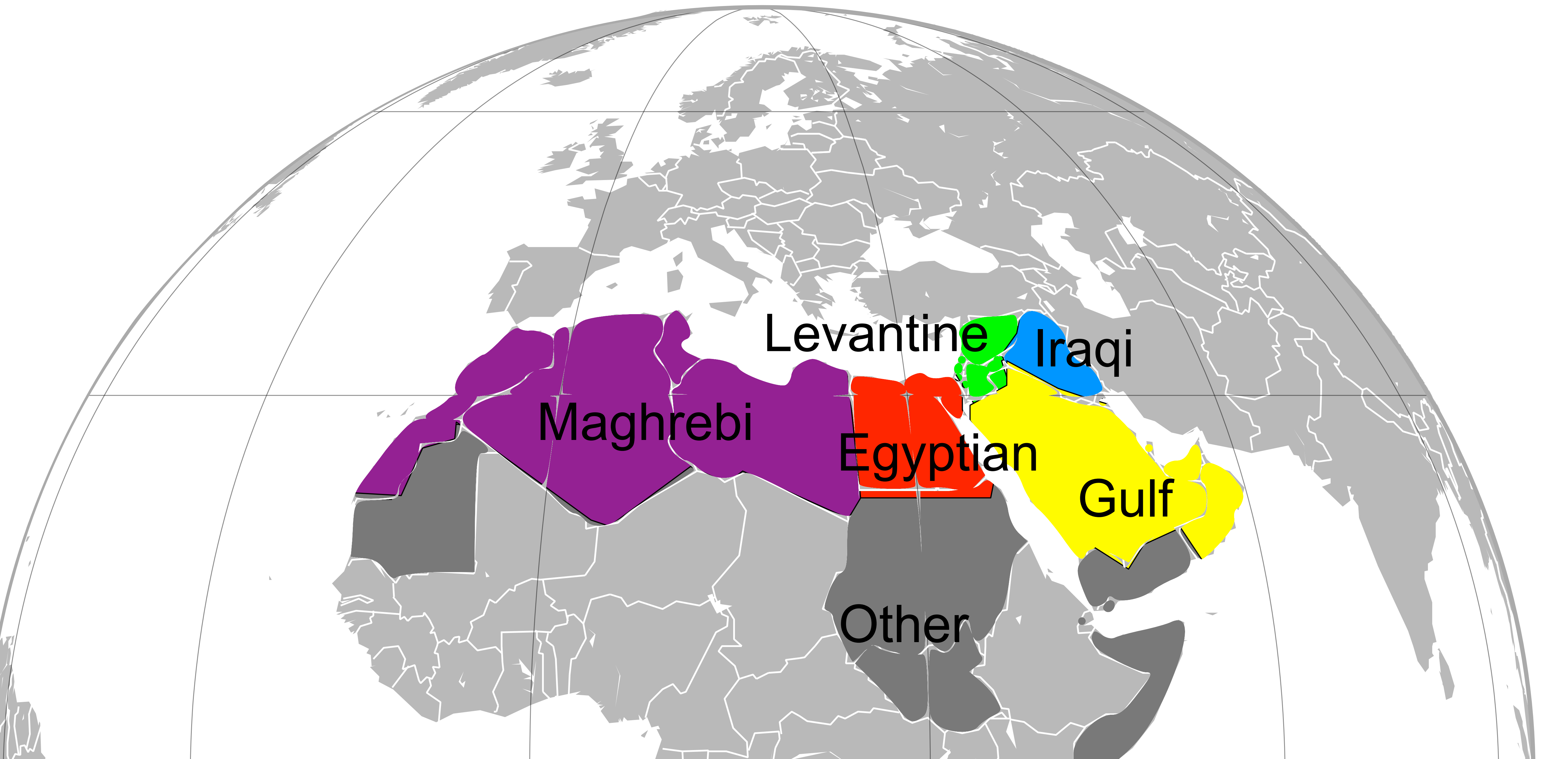
Professional Quality from Non-Professionals

Full details in Zaidan and Callison-Burch (ACL 2011a) & Zaidan (PhD Thesis 2012)



Gathering Data about Arabic Dialects

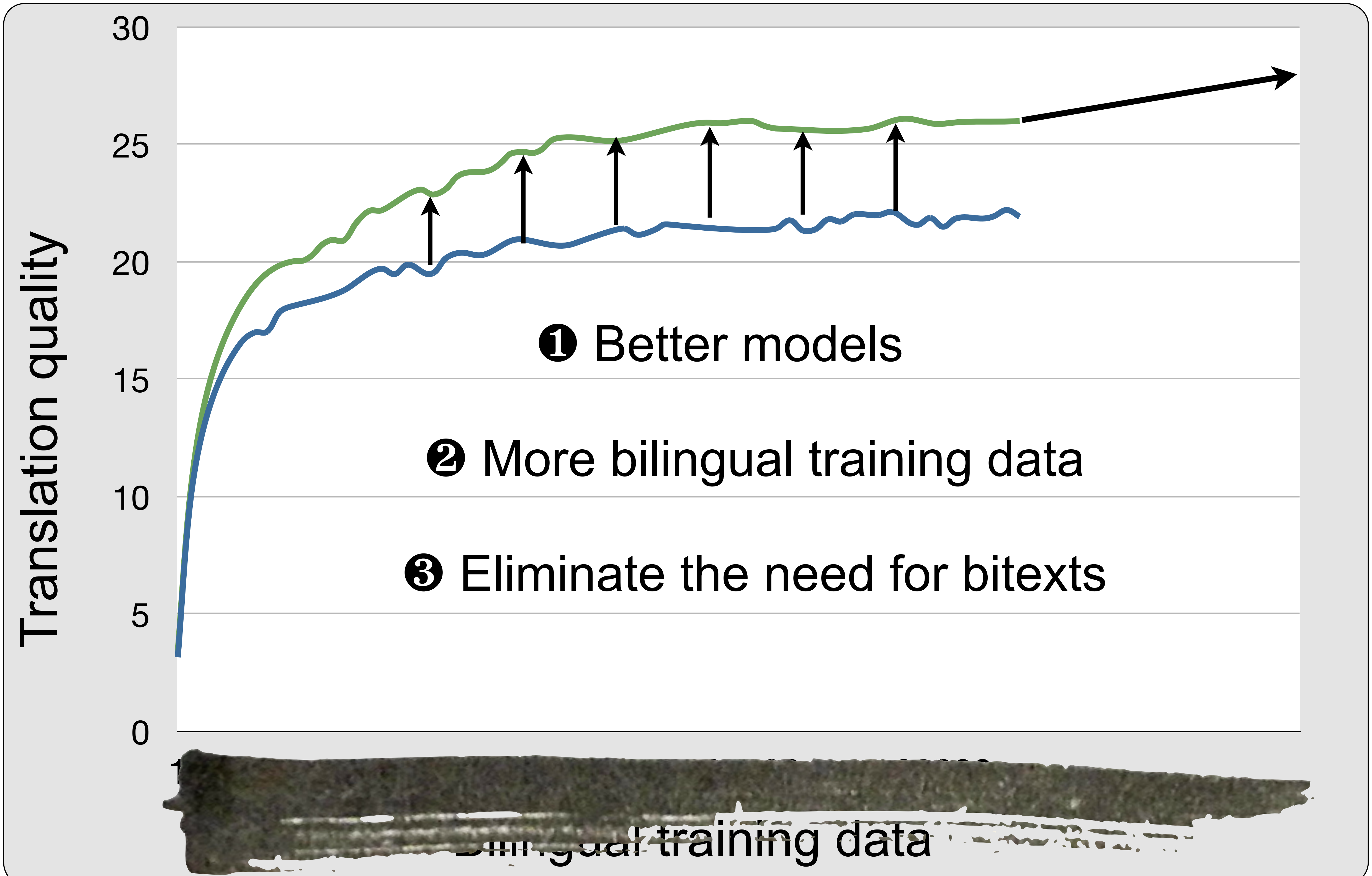
Arabic has different varieties. MSA is the standardized form but there are many distinct regional dialects.



Examples of Dialect Translation

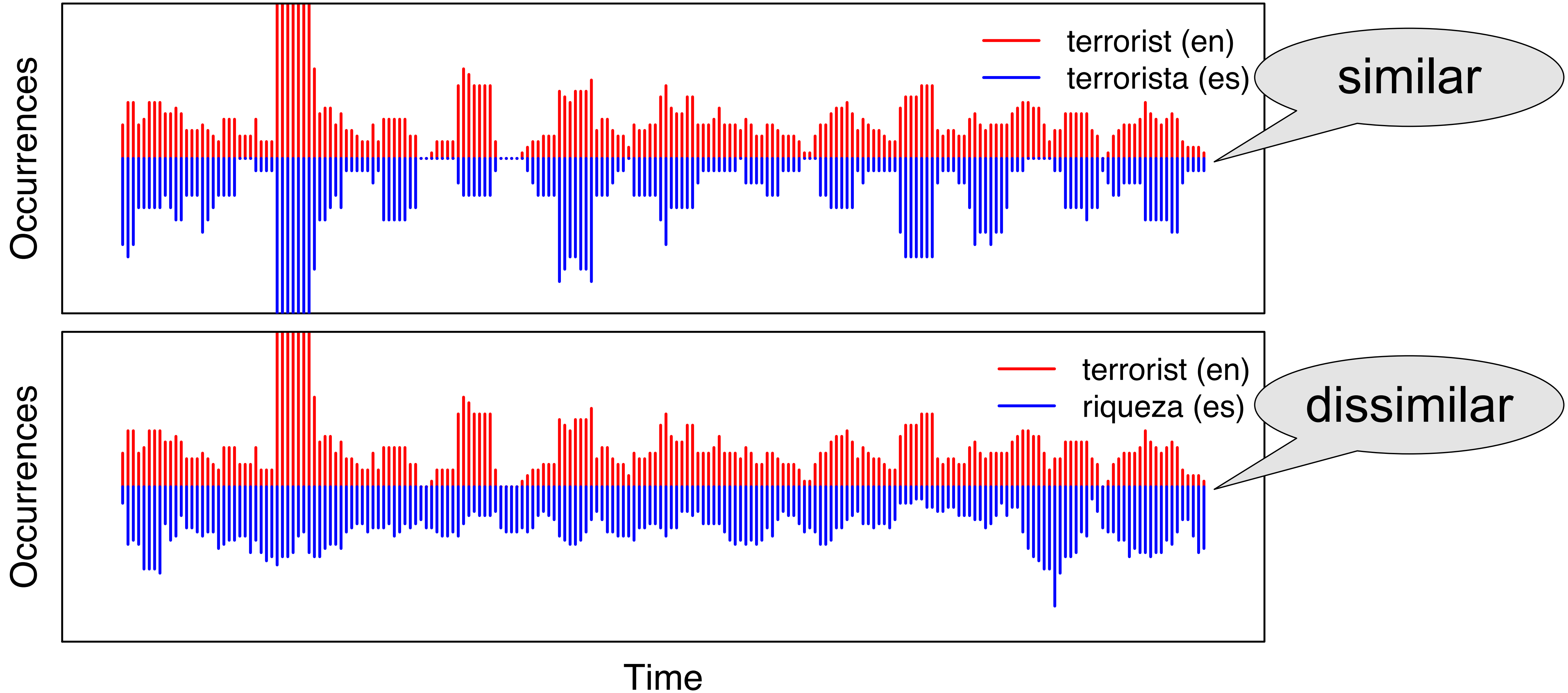
Dialect Input	MSA system	Dialect system	Reference
EGY انت بتعمــل لــه اعـــلان ولا ايــه ؟ !!	You are working for a declaration and not?	You are making the advertisement for him or what?	Are you promoting it or what?!!
EGY نفسي اطمئن عليه بعد ما شاف الصوره دي	Myself feel to see this image.	I wish to check on him after he saw this picture.	I want to be sure that he is fine after he saw the images.
LEV لهيك الجو كتيير كووول	God you the atmosphere.	This is why the weather is so cool	This is why the weather is so cool
LEV طول بالك عم نمزح	Do you think about a joke long.	Calm down we are kidding	Calm down, we are only kidding

How to Improve Machine Translation




reclamo otra vez cargos políticos
Fue una demostración de fuerza del aparato gremial. Pidió la reelección de Cristina. Pero insistió en que el sindicalismo tenga candidatos en las listas. Para la CGT, hubo 500 mil personas. Para la Policía y la SIDE, unas diez veces menos.
Cristina saludó por carta y envió a sus ministros
Boda Real
Como una película, que vieron dos mil millones
נשיקה
東亞日報
軍 "北 이상징후 포착" 경계강화
제3노총 6월에 출범
नवदुनिया
1900
200
30
4
प्रधानमंत्री को घेरा
तहसीलदार के सामने किसान ने खाया जहर
Las Últimas Noticias
tiroteo en que murió Osama bin Laden
2-6
מעריב
שם קוד: אלבים
כל הפרטים, מכל הזווית, דקה אחר דקה, של חיסול המבוקש מספר 1
בן מות
중앙일보
9·11 10년 '빈 라덴 추격' 끝내다
6년간 흑자 본식 대주주·경영진만 329억 배당 챙겨
বিন লাদেন
১০ বছরের লড়াই শেষ ৪০ মিনিটে
RIOdePERNAMBUCO
1955
2011
STEVE JOBS
O HOMEM QUE DEU ROSTO AO FUTURO
הארץ
האיש ששינה את חיינו
朝鮮日報
월街 시위에 美노조 가세
"70兆 날릴 각오" 삼성, 애플과 끝장승부
ब्रेकिंग न्यूज़:
एप्पल के करिश्माई सहसंस्थापक का कैंसर से निधन
जॉब्स दुनिया से अलविदा

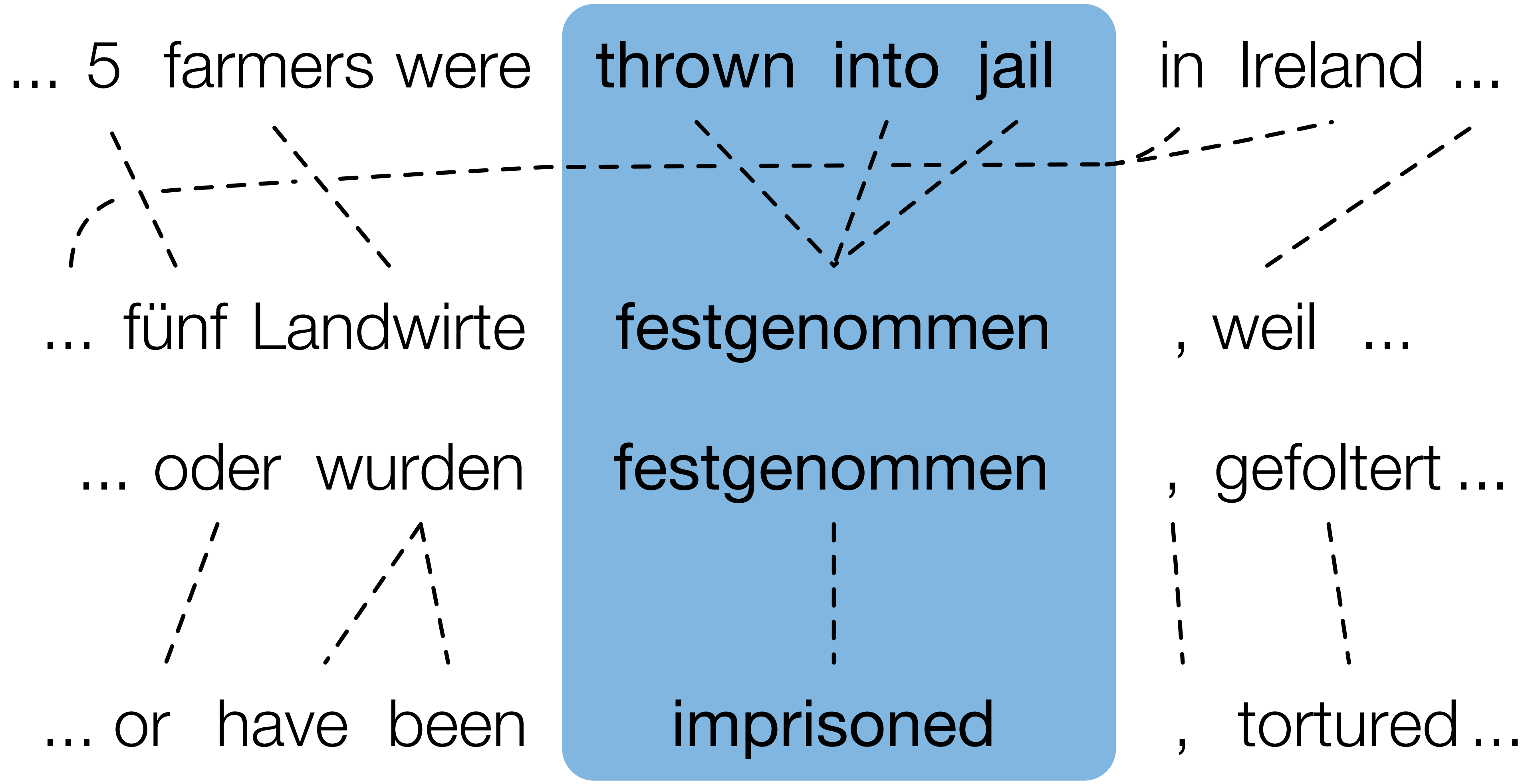
Scoring Translations: Time



Scoring Translations: Time

eólica	estambul	terrorista	vacuno
wind	**istanbul**	**terrorist**	beef
renewable	erdogan	terrorism	cattle
solar	turkish	terrorists	bse
sources	turkey	attacks	compulsory
renewables	turks	fight	meat
energy	ankara	attack	cows
energies	membership	terror	veal
electricity	negotiations	acts	**cow**
photovoltaic	undcp	threat	labelling
grid	talks	september	papayannakis

Paraphrasing

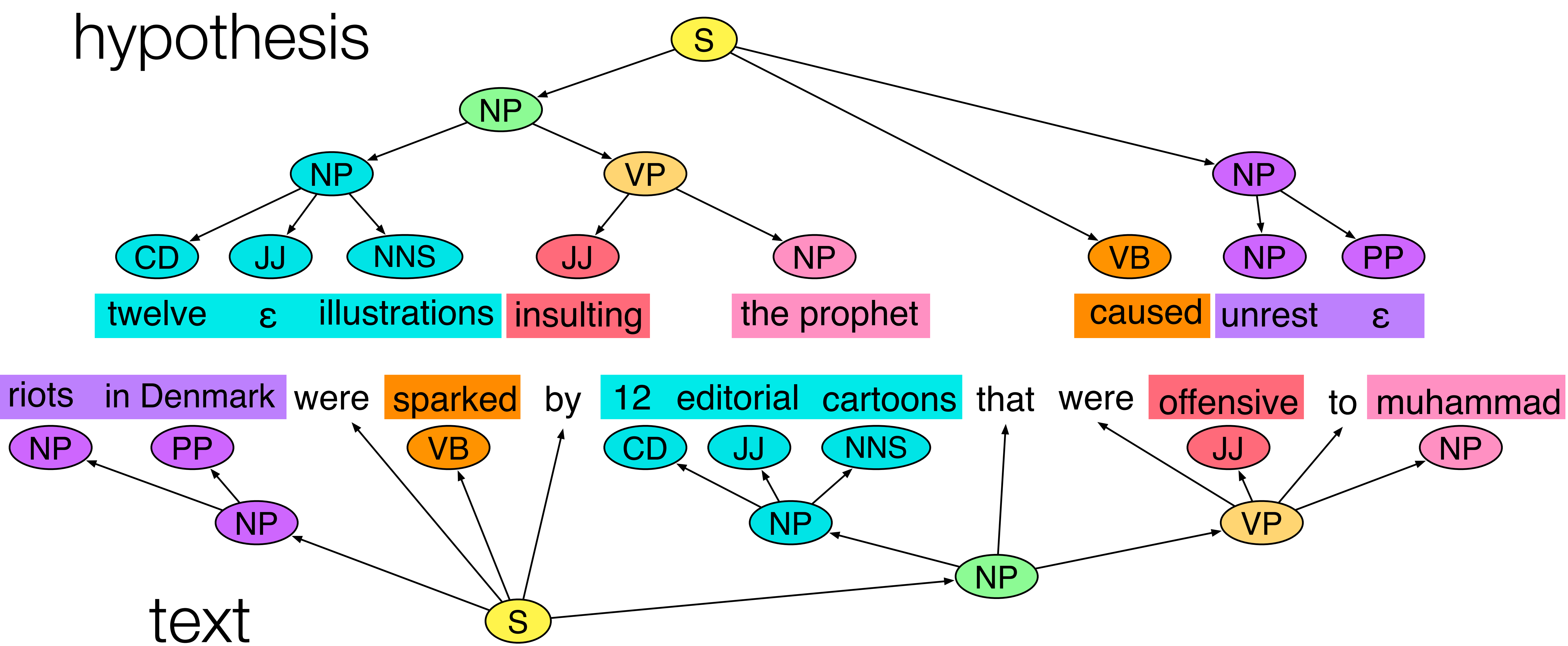


Many equivalent English expressions

thrown into jail

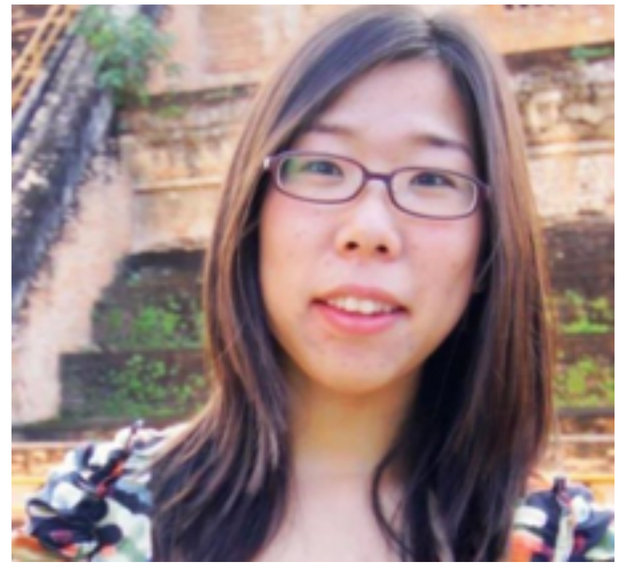
arrested	be thrown in prison	arrest
detained	been thrown into jail	cases
imprisoned	being arrested	custody
incarcerated	in jail	maltreated
jailed	in prison	owners
locked up	put in prison for	protection
taken into custody	were thrown into jail	thrown
thrown into prison	who are held in detention	

Natural Language Understanding



Guest Lecturers

Liang Huang
CUNY



Wei Xu
Penn

Matt Post
JHU

Will Lewis
MSR

Christian Buck
Edinburgh

Ken Heafield
Bloomberg

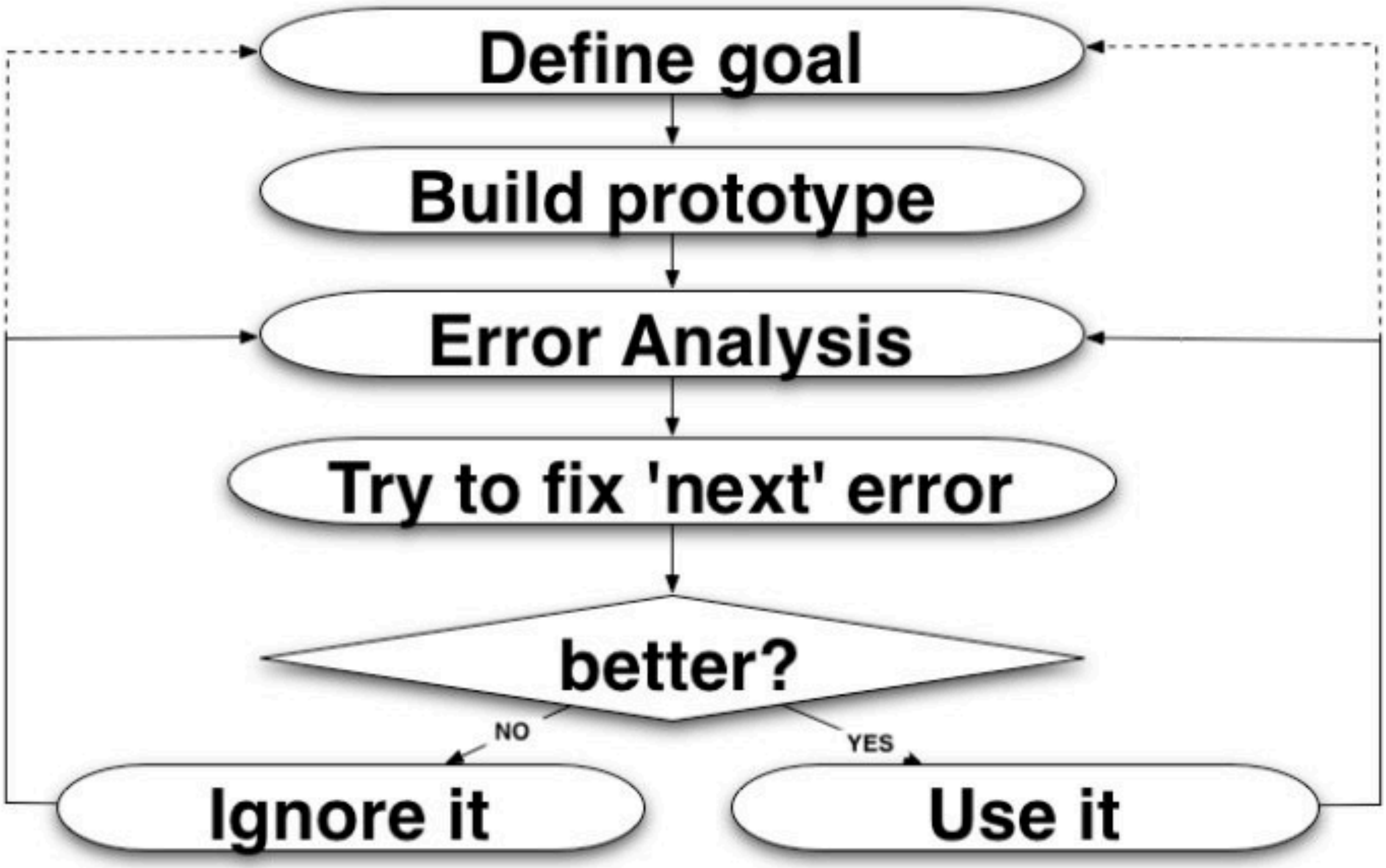
You guys!

Leaderboard

Alias	HW 0 # Correct	HW 1 AER	HW 2 Model Score ▲	HW 3 Accuracy	HW 4 BLEU
okyurp	10	15.12	-1211.39	0.548676	28.3847
do_not_set_yourself_on_fire	10	18.73	-1224.20	0.540384	27.3337
fly	10	26.94	-1228.65	0.544452	27.9741
@jim	10	16.59	-1229.52	0.547972	27.6029
sogeking	10	18.02	-1230.32	0.556538	28.3220
lilies	10	14.73	-1237.72	0.560918	27.5001
direKt translation	10	21.81	-1238.51	0.541636	28.1127
StopItRon	10	34.04	-1239.82	0.532835	28.1132
Cloud9	10	27.09	-1242.32	0.544687	28.6879
Etaoin Shrdlu	10	19.03	-1242.62	0.539797	28.5747
☃	10	26.81	-1243.78	0.533657	27.8708
mstag	10	31.41	-1244.77	0.527281	27.5254
Dhrubeel	10	27.59	-1245.83	0.562209	27.9579
Class the MT is.	10	28.32	-1246.62	0.544256	28.5456
Aafikins	10	27.59	-1249.20	0.561818	28.4572
aoc	10	20.51	-1254.07	0.553448	27.4141
sqq	10	25.45	-1254.24	0.555912	28.0958
Kailoofi	10	25.76	-1257.08	0.557868	28.4231
ohayyy	10	26.53	-1263.84	0.538076	27.2392
John E. Mason	10	26.48	-1269.91	0.540345	28.6509
Baby Eigensheep	10	30.60	-1270.60	0.550045	27.0876
Mithrandir	10	27.51	-1274.59	0.560488	28.2648
Chief Relief	10	27.19	-1279.60	0.556616	27.3528
toffl	10	27.40	-1283.94	0.528533	28.2981
luck	10	26.70	-1295.9[illegible]	0.541714	27.1022
mogjuice	10	31.10	-1298.83	0.540462	27.8330
Madan-Mohan Das	10	23.22	-1299.82	0.537333	23.6436

Moses
(off the shelf):
-1286.92

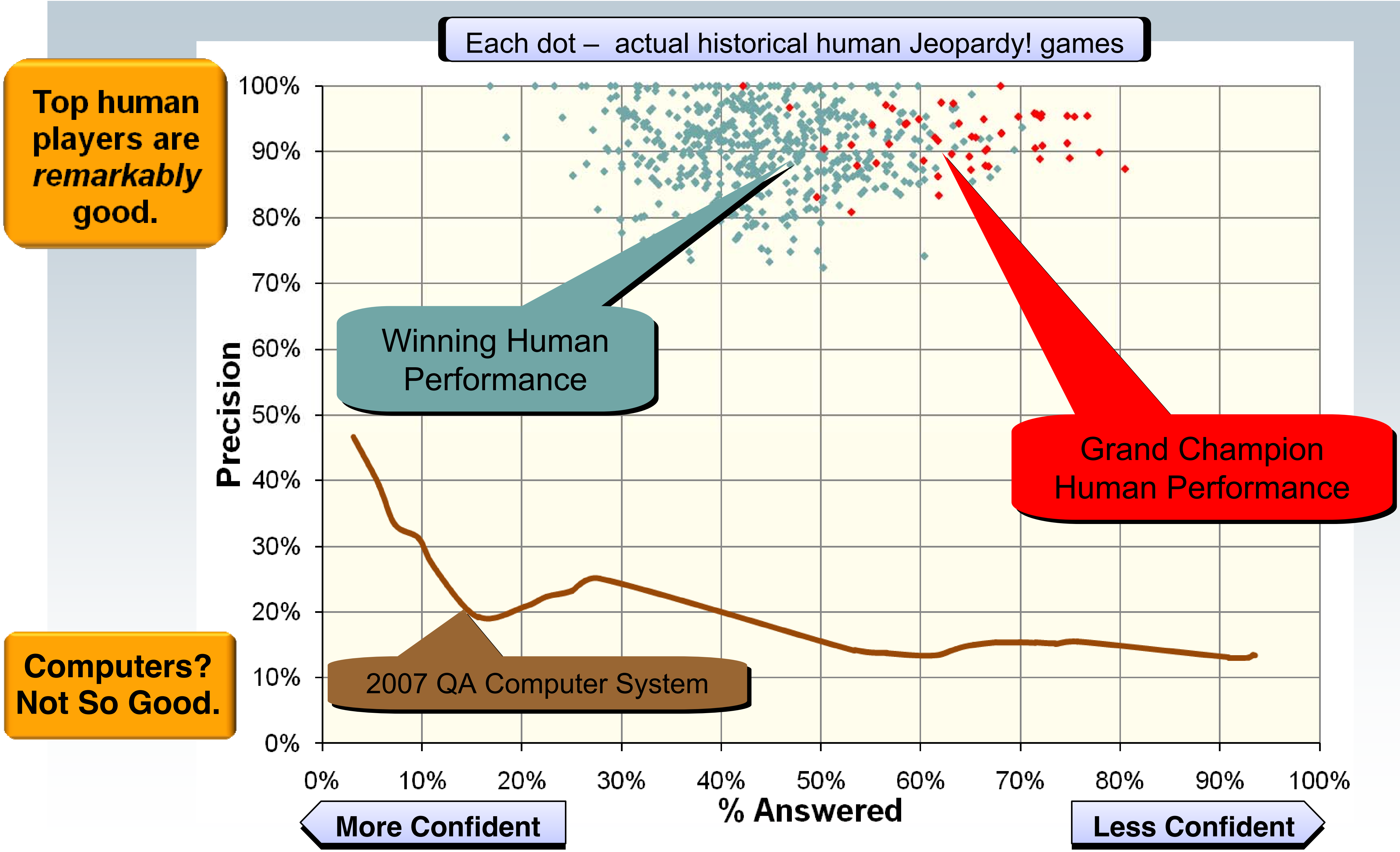
Development Cycle for MT Research



Google

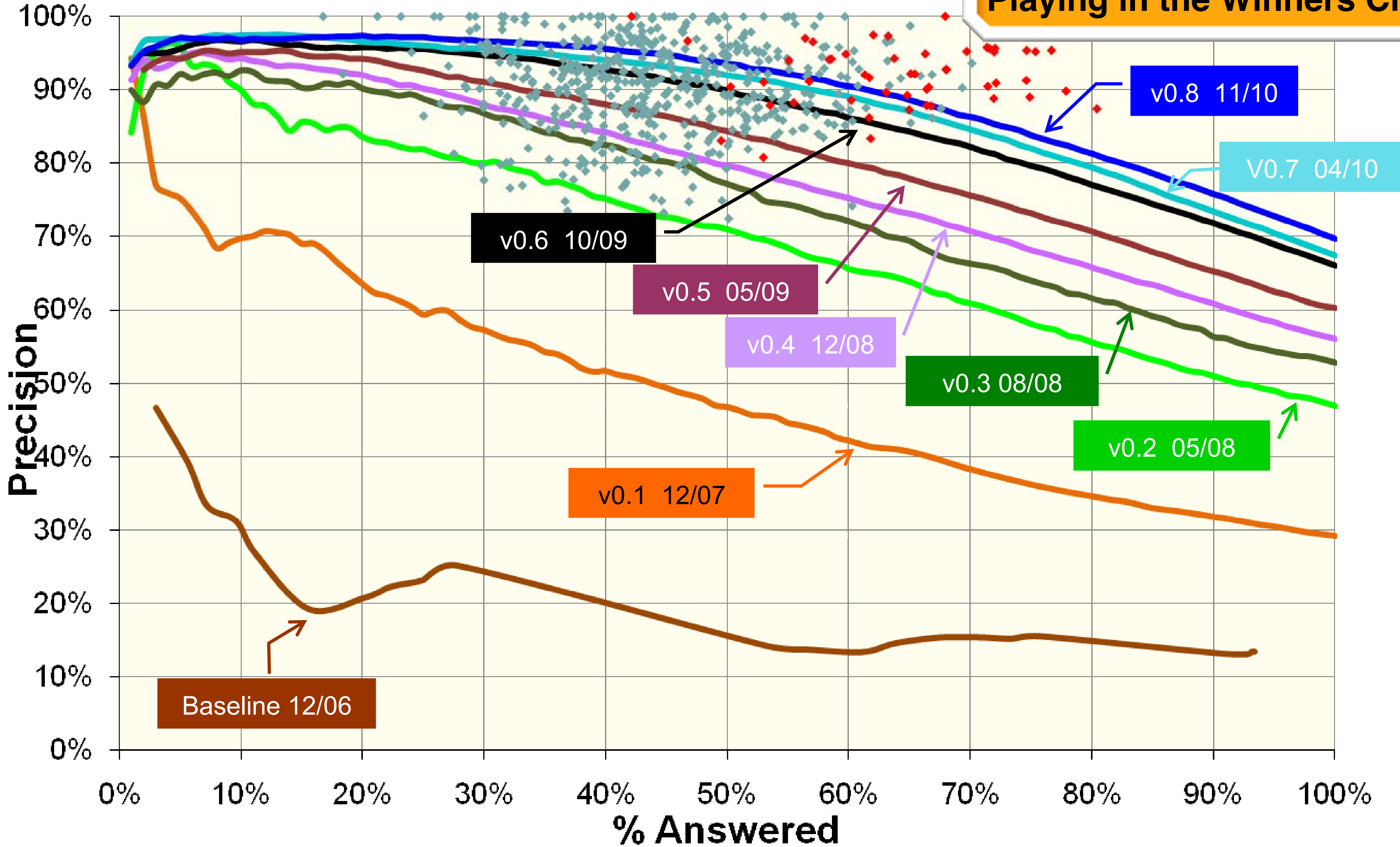
What It Takes to compete against Top Human Jeopardy! Players

*Our Analysis Reveals the **Winner's Cloud***



DeepQA: Incremental Progress in Answering Precision on the Jeopardy Challenge: 6/2007-11/2010

IBM *Watson*
Playing in the Winners Cloud



Your term projects did this.

- You defined a challenge problem
- You defined a scoring function that allowed you to plot your progress over time.
- You and your classmates tried different algorithms and developed different models to solve the problem.

What you can do to stay involved

- If you will be back at Penn next year:
  - Take CIS 530 - Computation Linguistics
  - Take CIS 520 - Machine Learning
  - Do an independent research project with me!
- If you're graduating, then stay in touch!

Thanks!

Stay in touch!

email: ccb@cis.upenn.edu
Twitter: @ccb
Facebook: chris.callisonburch