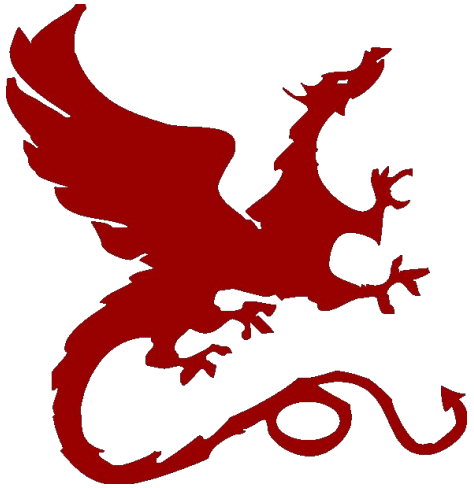
Algorithms for NLP

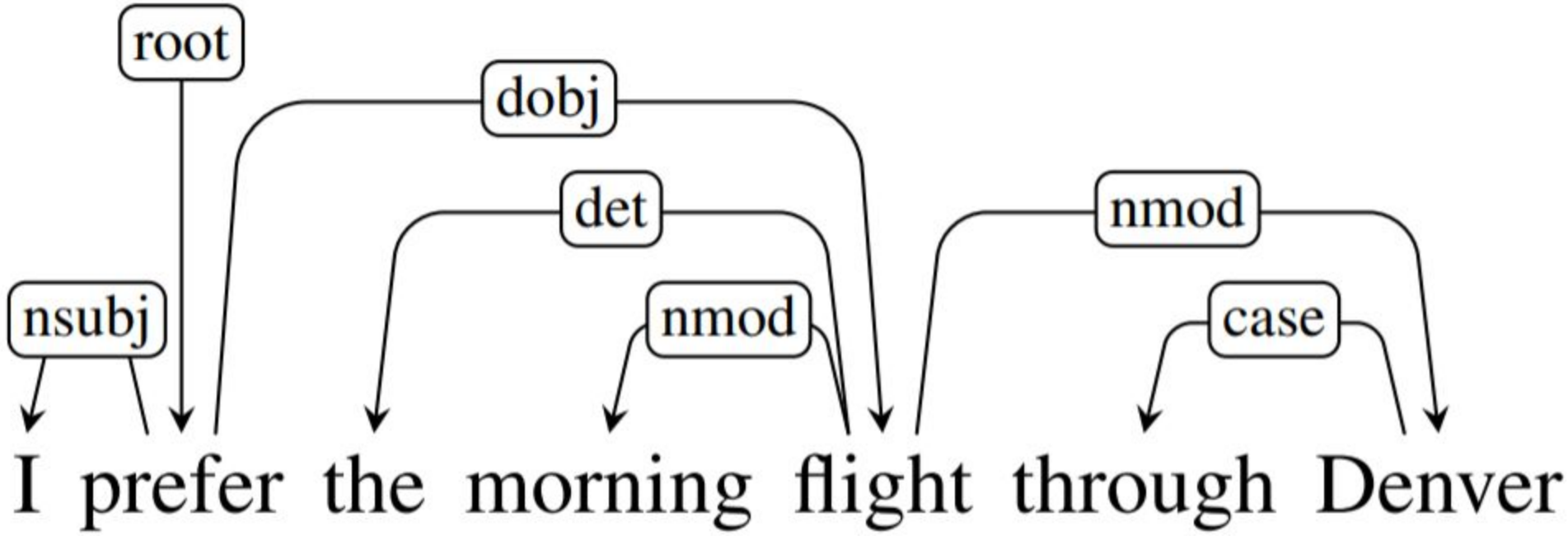


Machine Translation I

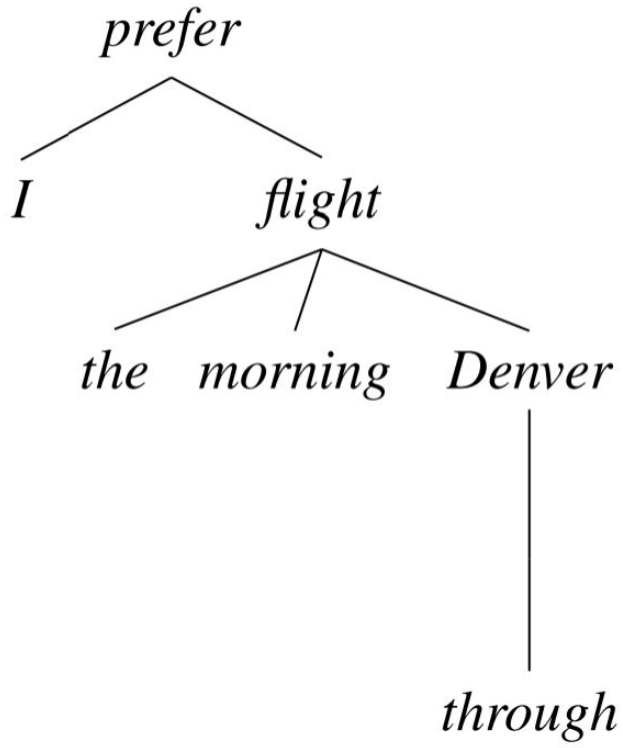
Yulia Tsvetkov – CMU

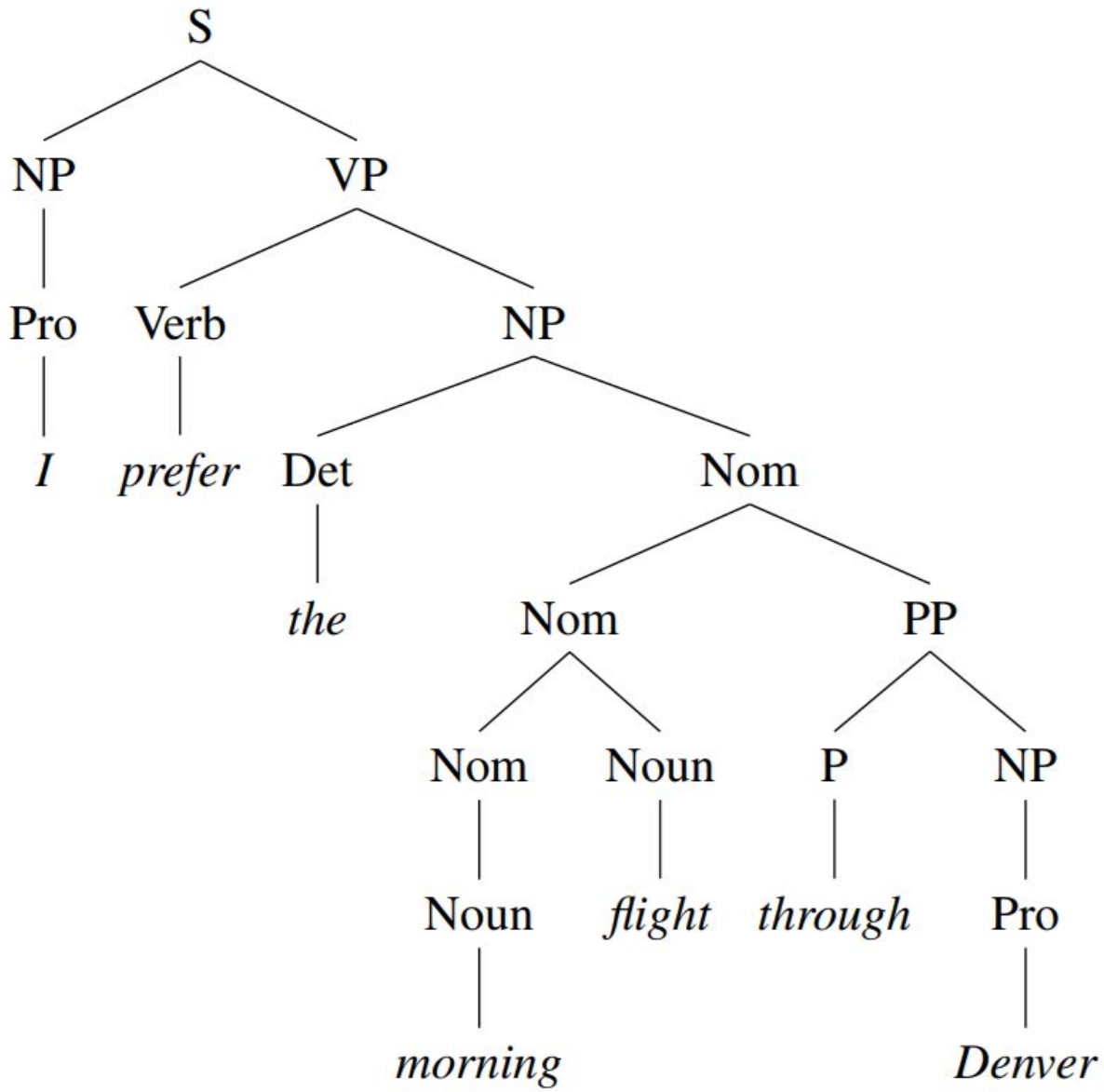
Slides: Chris Dyer – DeepMind;
Taylor Berg-Kirkpatrick – CMU/UCSD, Dan Klein – UC Berkeley

Dependency representation



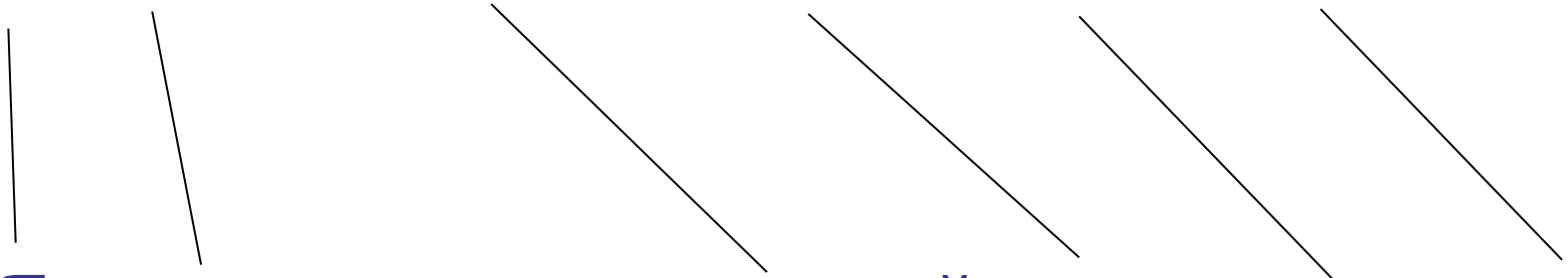
Dependency vs Constituency trees





Languages with free word order

I prefer the morning flight through Denver



Я предпочитаю утренний перелет через Денвер

Я предпочитаю через Денвер утренний перелет

Утренний перелет я предпочитаю через Денвер

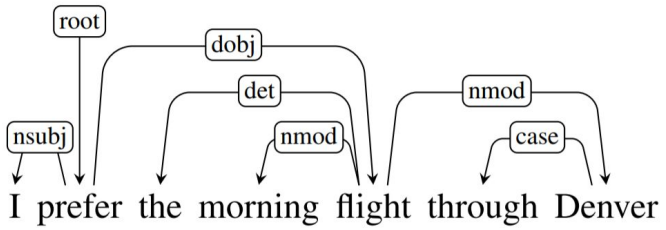
Перелет утренний я предпочитаю через Денвер

Через Денвер я предпочитаю утренний перелет

Я через Денвер предпочитаю утренний перелет

…

Dependency Constraints

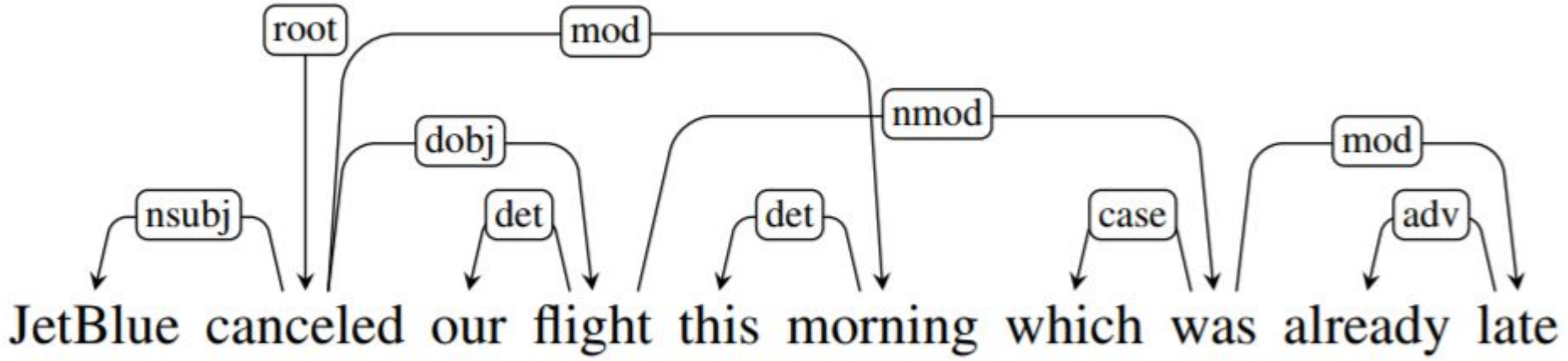


- Syntactic structure is complete (connectedness)
  - connectedness can be enforced by adding a special root node
- Syntactic structure is hierarchical (acyclicity)
  - there is a unique pass from the root to each vertex
- Every word has at most one syntactic head (single-head constraint)
  - except root that does not have incoming arcs

This makes the dependencies a tree

Projectivity

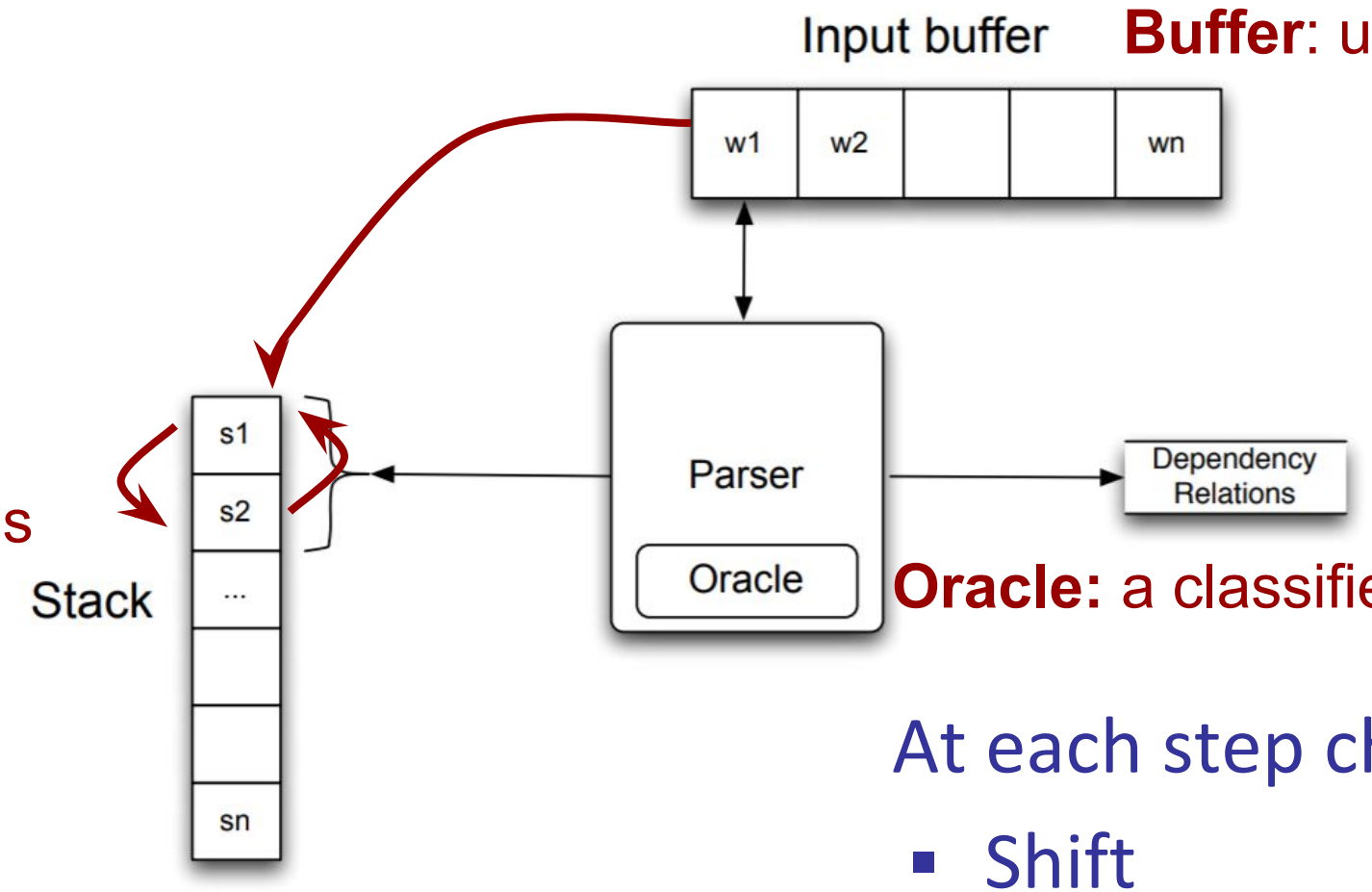
- Projective parse
  - arcs don’t cross each other
  - mostly true for English
- Non-projective structures are needed to account for
  - long-distance dependencies
  - flexible word order



Parsing algorithms

- Transition based
  - greedy choice of local transitions guided by a goodclassifier
  - deterministic
  - MaltParser (Nivre et al. 2008)
- Graph based
  - Minimum Spanning Tree for a sentence
  - McDonald et al.'s (2005) MSTParser
  - Martins et al.'s (2009) Turbo Parser

Configuration for transition-based parsing



At each step choose:

- Shift
- LeftArc or Reduce left
- RightArc or Reduce right

Shift-Reduce Parsing

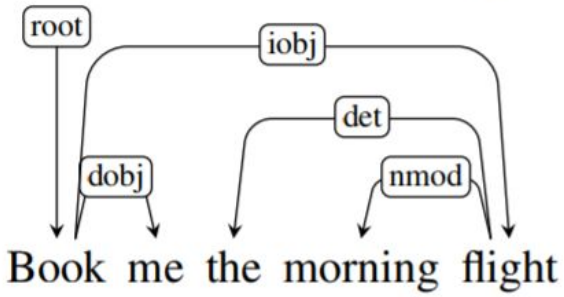
Configuration:

- Stack, Buffer, Oracle, Set of dependency relations

Operations by a classifier at each step:

- Shift
  - remove w1 from the buffer, add it to the top of the stack as s1
- LeftArc or Reduce left
  - assert a head-dependent relation between s1 and s2
  - remove s2 from the stack
- RightArc or Reduce right
  - assert a head-dependent relation between s2 and s1
  - remove s1 from the stack

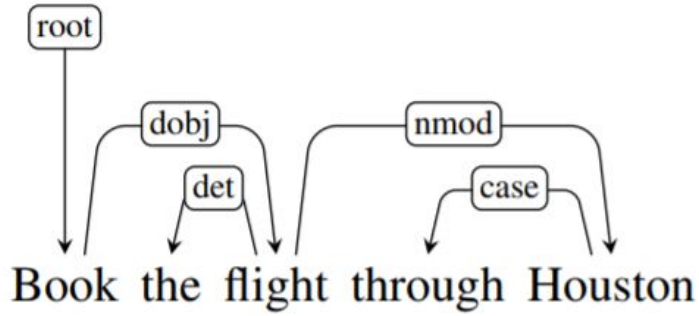
Shift-Reduce Parsing (arc-standard)



Step	Stack	Word List	Action	Relation Added
0	[root]	[book, me, the, morning, flight]	SHIFT	
1	[root, book]	[me, the, morning, flight]	SHIFT	
2	[root, book, me]	[the, morning, flight]	RIGHTARC	(book → me)
3	[root, book]	[the, morning, flight]	SHIFT	
4	[root, book, the]	[morning, flight]	SHIFT	
5	[root, book, the, morning]	[flight]	SHIFT	
6	[root, book, the, morning, flight]	[]	LEFTARC	(morning ← flight)
7	[root, book, the, flight]	[]	LEFTARC	(the ← flight)
8	[root, book, flight]	[]	RIGHTARC	(book → flight)
9	[root, book]	[]	RIGHTARC	(root → book)
10	[root]	[]	Done	

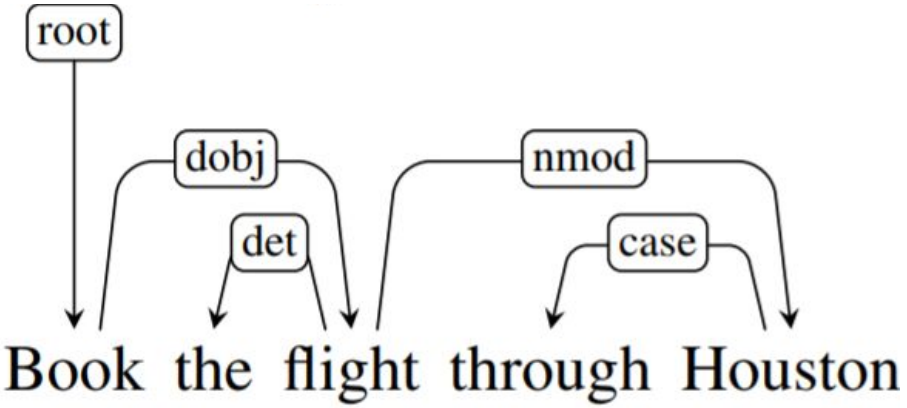
Training an Oracle

- How to extract the training set?
  - if LeftArc → LeftArc
  - if RightArc
    - if s1 dependents have been processed → RightArc
  - else → Shift



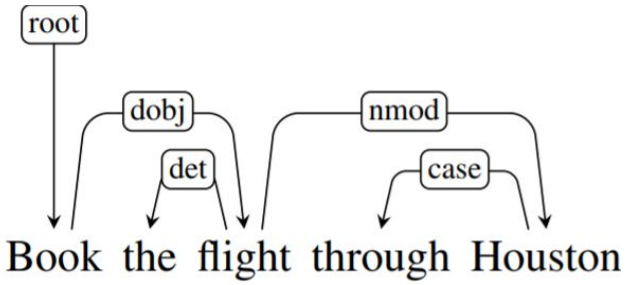
Step	Stack	Word List	Predicted Action
0	[root]	[book, the, flight, through, houston]	SHIFT
1	[root, book]	[the, flight, through, houston]	SHIFT
2	[root, book, the]	[flight, through, houston]	SHIFT
3	[root, book, the, flight]	[through, houston]	LEFTARC
4	[root, book, flight]	[through, houston]	SHIFT
5	[root, book, flight, through]	[houston]	SHIFT
6	[root, book, flight, through, houston]	[]	LEFTARC
7	[root, book, flight, houston]	[]	RIGHTARC
8	[root, book, flight]	[]	RIGHTARC
9	[root, book]	[]	RIGHTARC

Arc-Eager



- LEFTARC: Assert a head-dependent relation between s1 and b1; pop the stack.
- RIGHTARC: Assert a head-dependent relation between s1 and b1; shift b1 to be s1.
- SHIFT: Remove b1 and push it to be s1.
- REDUCE: Pop the stack.

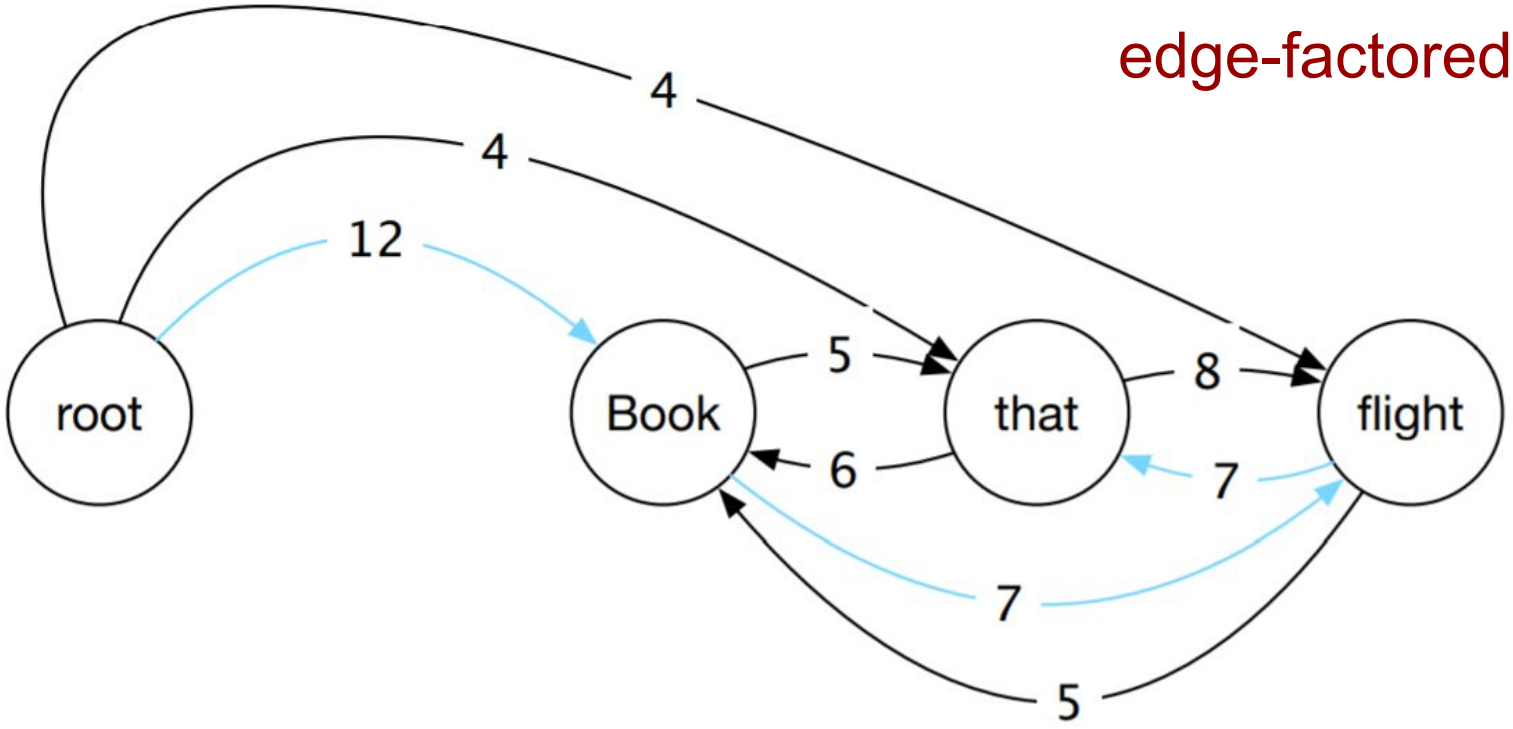
Arc-Eager



Step	Stack	Word List	Action	Relation Added
0	[root]	[book, the, flight, through, houston]	RIGHTARC	(root → book)
1	[root, book]	[the, flight, through, houston]	SHIFT	
2	[root, book, the]	[flight, through, houston]	LEFTARC	(the ← flight)
3	[root, book]	[flight, through, houston]	RIGHTARC	(book → flight)
4	[root, book, flight]	[through, houston]	SHIFT	
5	[root, book, flight, through]	[houston]	LEFTARC	(through ← houston)
6	[root, book, flight]	[houston]	RIGHTARC	(flight → houston)
7	[root, book, flight, houston]	[]	REDUCE	
8	[root, book, flight]	[]	REDUCE	
9	[root, book]	[]	REDUCE	
10	[root]	[]	Done	

Graph-Based Parsing Algorithms

edge-factored approaches



- Start with a fully-connected directed graph
- Find a Minimum Spanning Tree
  - Chu and Liu (1965) and Edmonds (1967) algorithm

Chu-Liu Edmonds algorithm

function MAXSPANNINGTREE($G=(V,E)$, *root*, *score*) **returns** *spanning tree*

$F \leftarrow []$
$T' \leftarrow []$
score' $\leftarrow []$
for each $v \in V$ **do**
  bestInEdge $\leftarrow \text{argmax}_{e=(u,v) \in E}$ *score*[*e*]
  $F \leftarrow F \cup$ *bestInEdge*
  for each $e=(u,v) \in E$ **do**
    score'[*e*] $\leftarrow$ *score*[*e*] $-$ *score*[*bestInEdge*]
if $T=(V,F)$ is a spanning tree **then return** it
else
  $C \leftarrow$ a cycle in F
  $G' \leftarrow$ CONTRACT(G, C)
  $T' \leftarrow$ MAXSPANNINGTREE(G', *root*, *score'*)
  T $\leftarrow$ EXPAND(T', C)
  return T

function CONTRACT(G, C) **returns** *contracted graph*

function EXPAND(T, C) **returns** *expanded graph*

Select best incoming edge for each node

Subtract its score from all incoming edges

Stopping condition

Contract nodes if there are cycles

Recursively compute MST

Expand contracted nodes

Summary

- Transition-based
  - + Fast
  - + Rich features of context
  - - Greedy decoding
- Graph-based
  - + Exact or close to exact decoding
  - - Weaker features

Well-engineered versions of the approaches achieve comparable accuracy (on English), but make different errors

→ combining the strategies results in a substantial boost in performance

End of Previous Lecture

Machine Translation

Two Views of MT

- Direct modeling (aka pattern matching)
  - I have really good learning algorithms and a bunch of example inputs (source language sentences) and outputs (target language translations)

- Code breaking (aka the noisy channel, Bayes rule)
  - I know the target language
  - I have example translations texts (example enciphered data)

MT as Direct Modeling

$$\hat{e} = \arg\max_{e} p_{\lambda}(e \mid f)$$

target source

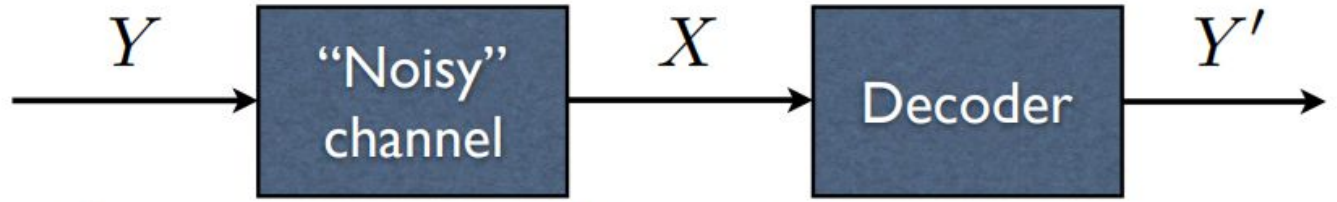
- one model does everything
- trained to reproduce a corpus of translations

MT as Code Breaking

One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: '***This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.***'

Warren Weaver to Norbert Wiener, March, 1947

Noisy Channel Model



Sent transmission

English

Received transmission

"French"

Recovered message

English'

$$\hat{e} = \arg\max_{e} p_{\varphi}(e) \times p_{\theta}(f \mid e)$$

language model

translation model

Which is better?

- Noisy channel - $p_{\theta}(e) \times p_{\varphi}(f \mid e)$
  - easy to use monolingual target language data
  - search happens under a product of two models (individual models can be simple, product can be powerful)
  - obtaining probabilities requires renormalizing

- Direct model - $p_{\lambda}(e \mid f)$
  - directly model the process you care about
  - model must be very powerful

Where are we in 2018?

- Direct modeling is where most of the action is
  - Neural networks are very good at generalizing and conceptually very simple
  - Inference in “product of two models” is hard

- Noisy channel ideas are incredibly important and still play a big role in how we think about translation

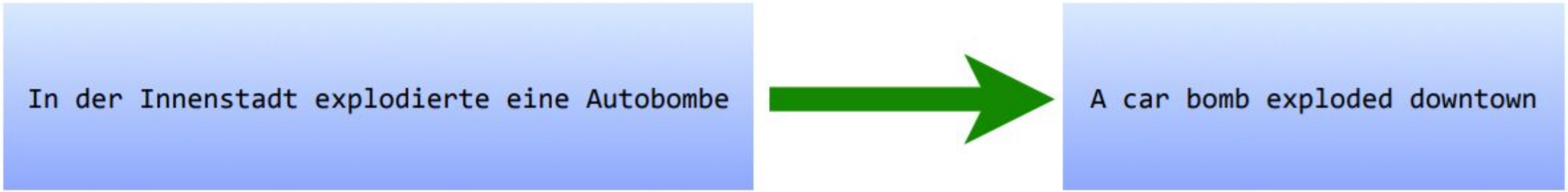
A common problem

$$\hat{e} = \arg\max_{e} p_{\varphi}(e) \times p_{\theta}(f \mid e)$$ Noisy channel

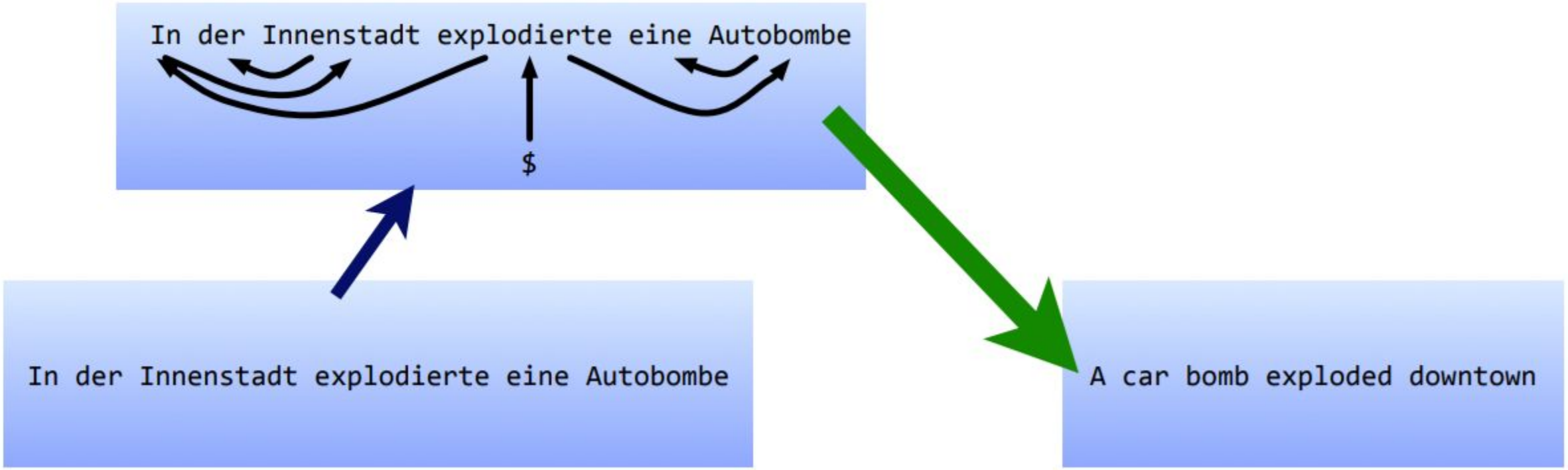
$$\hat{e} = \arg\max_{e} p_{\lambda}(e \mid f)$$ Direct

Both models must assign probabilities to how a sentence in one language translates into a sentence in another language.

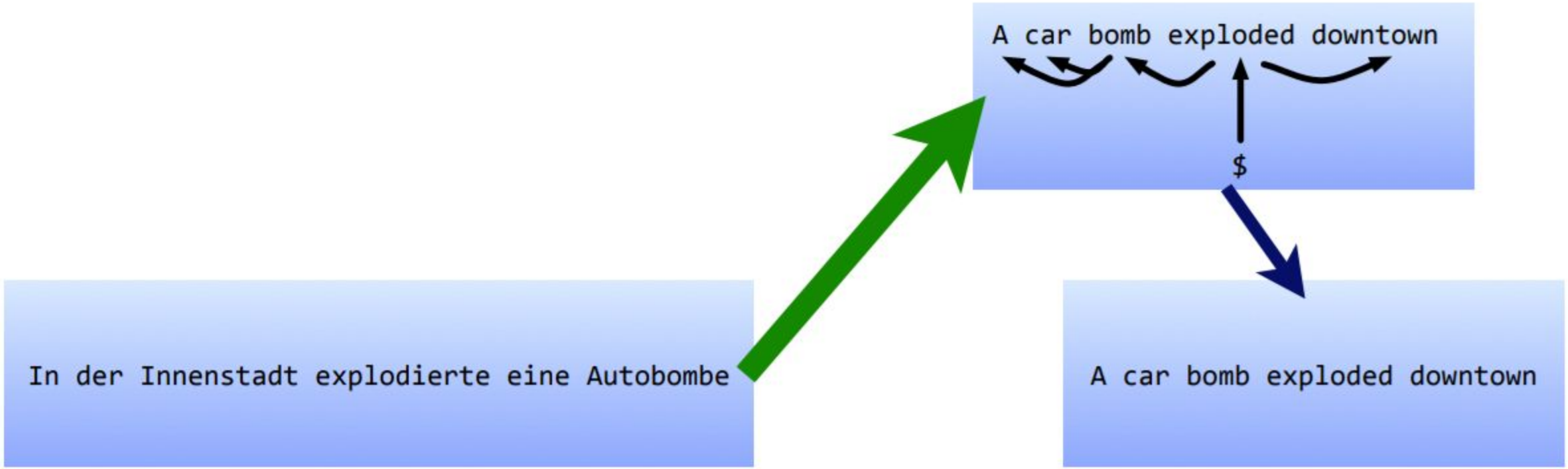
Levels of Transfer



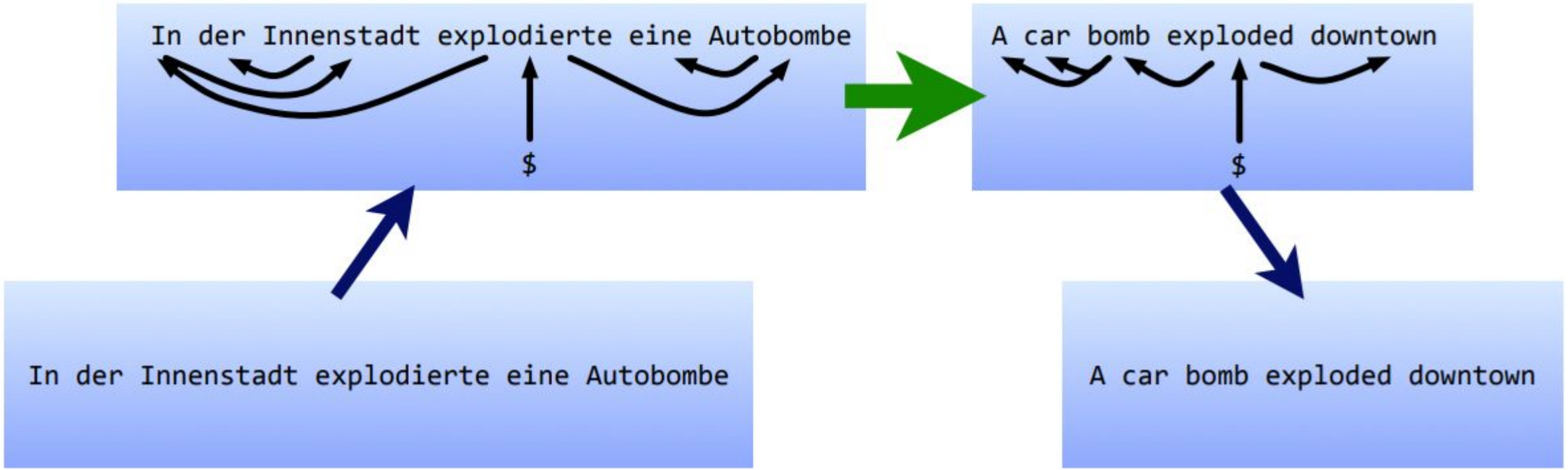
Levels of Transfer



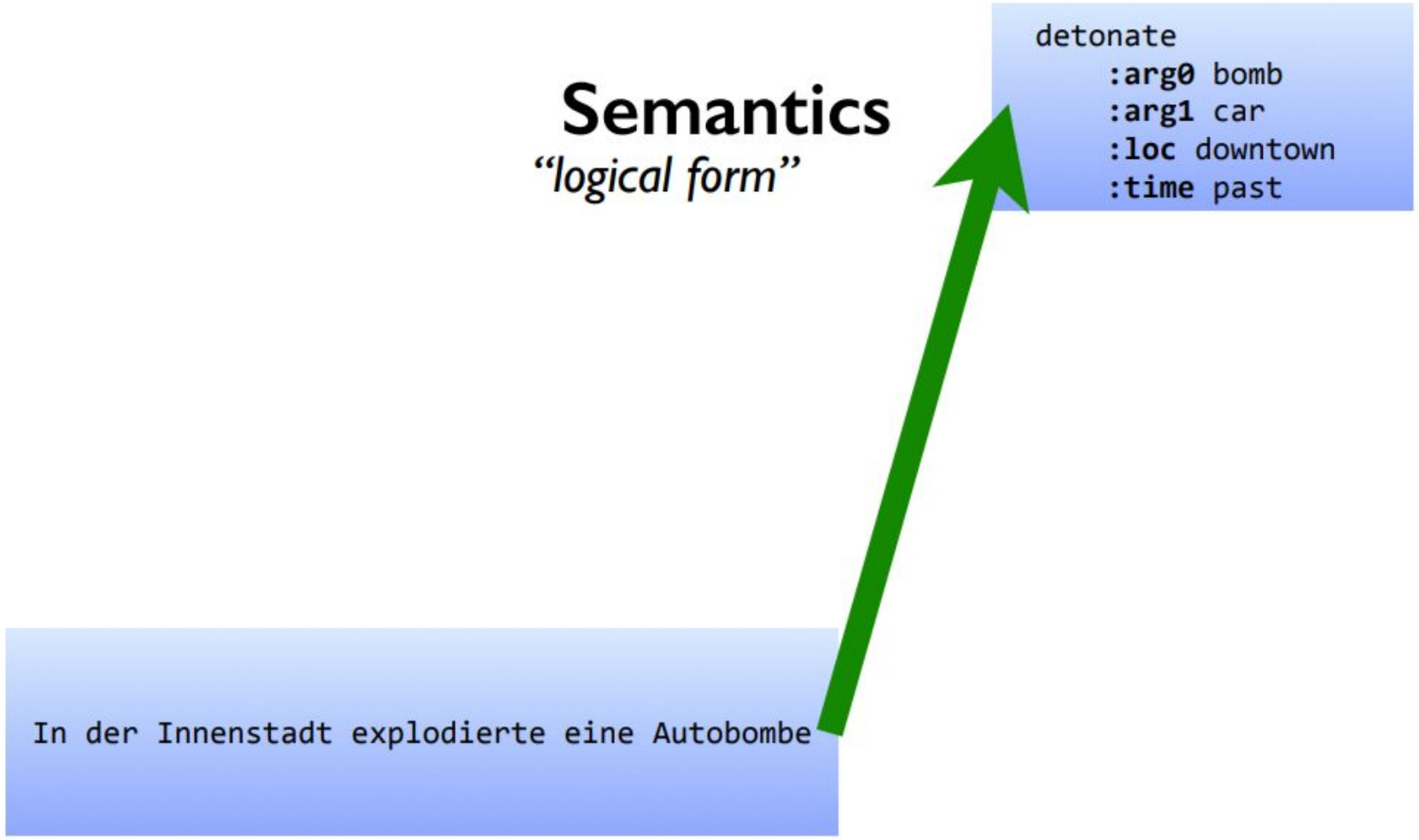
Levels of Transfer



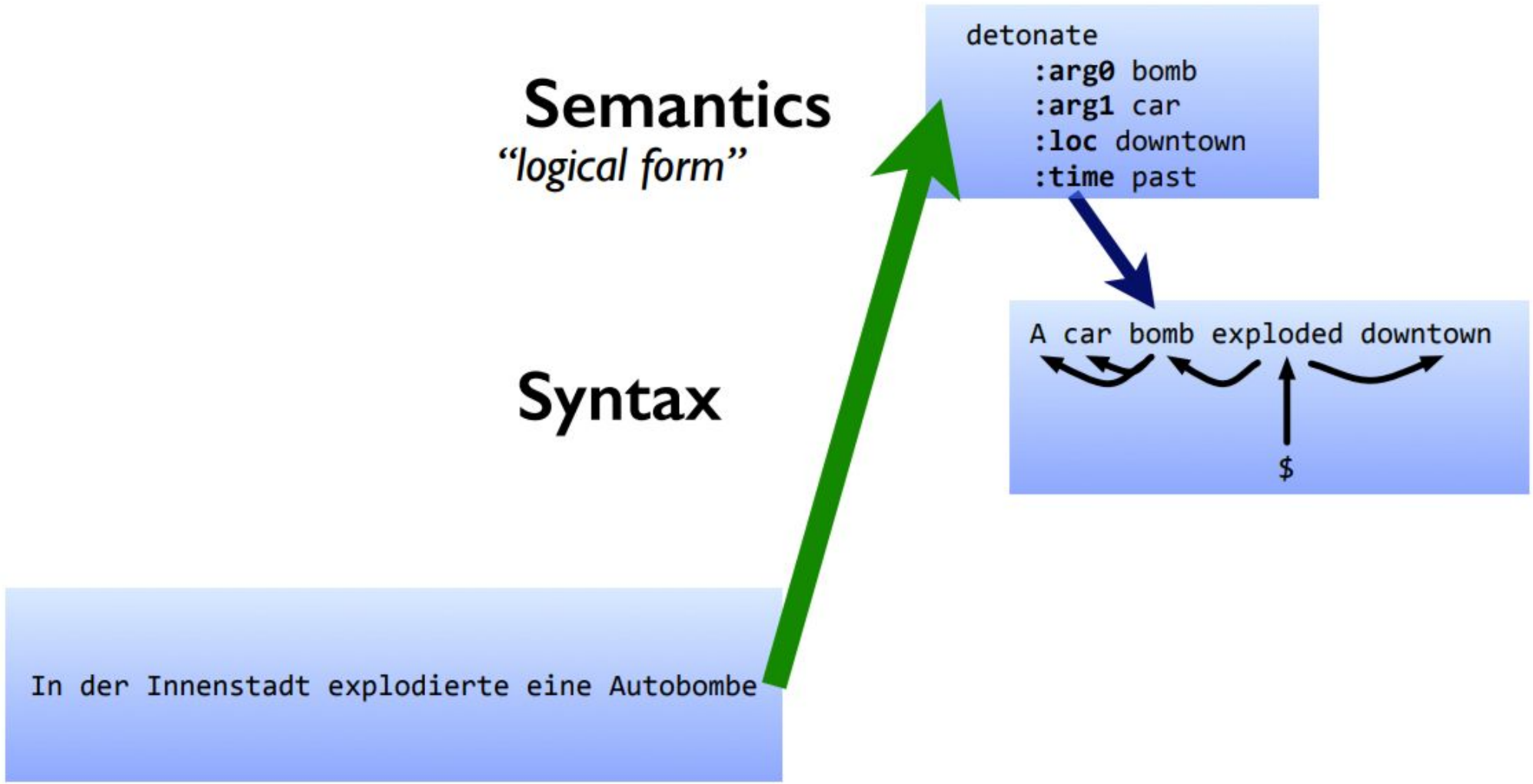
Levels of Transfer



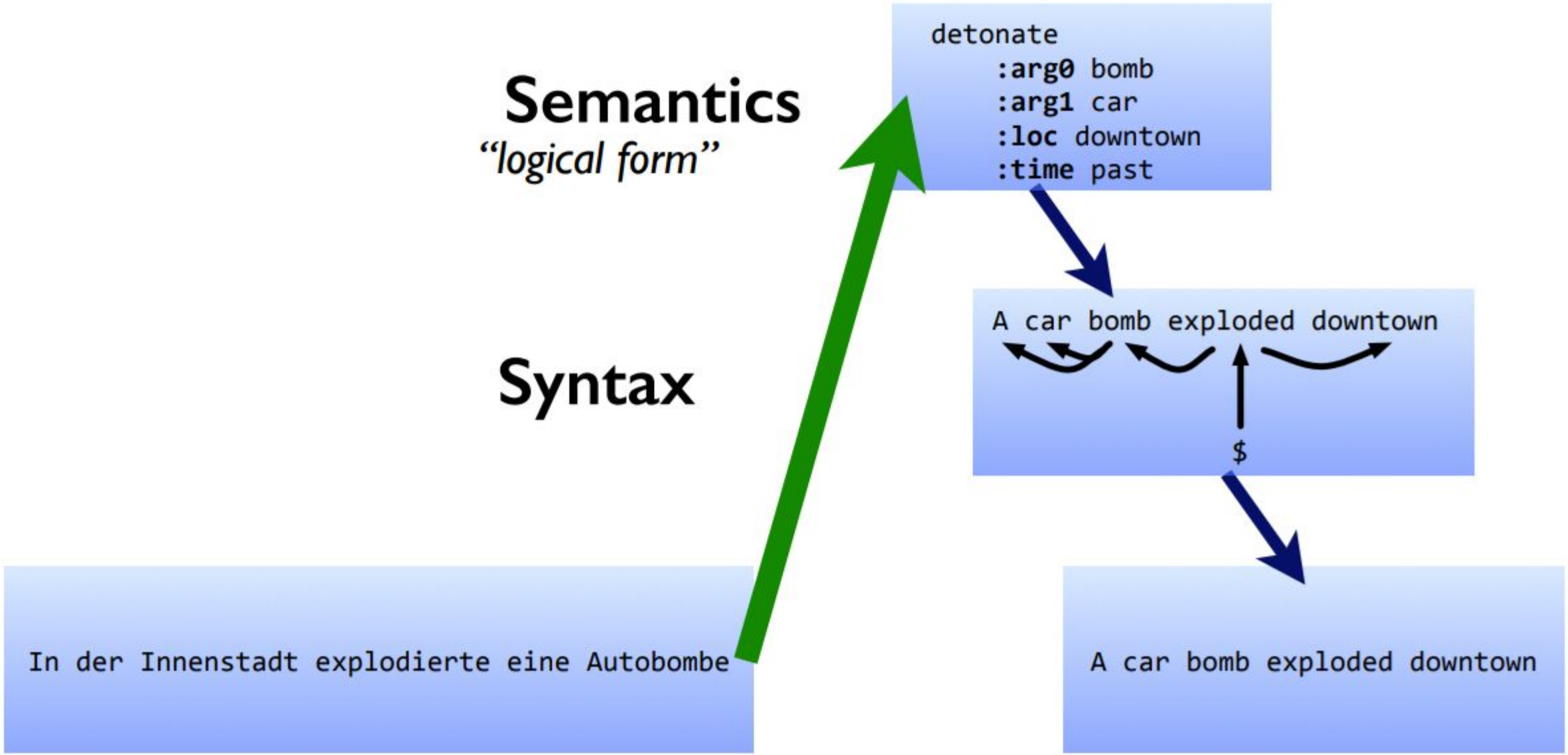
Levels of Transfer



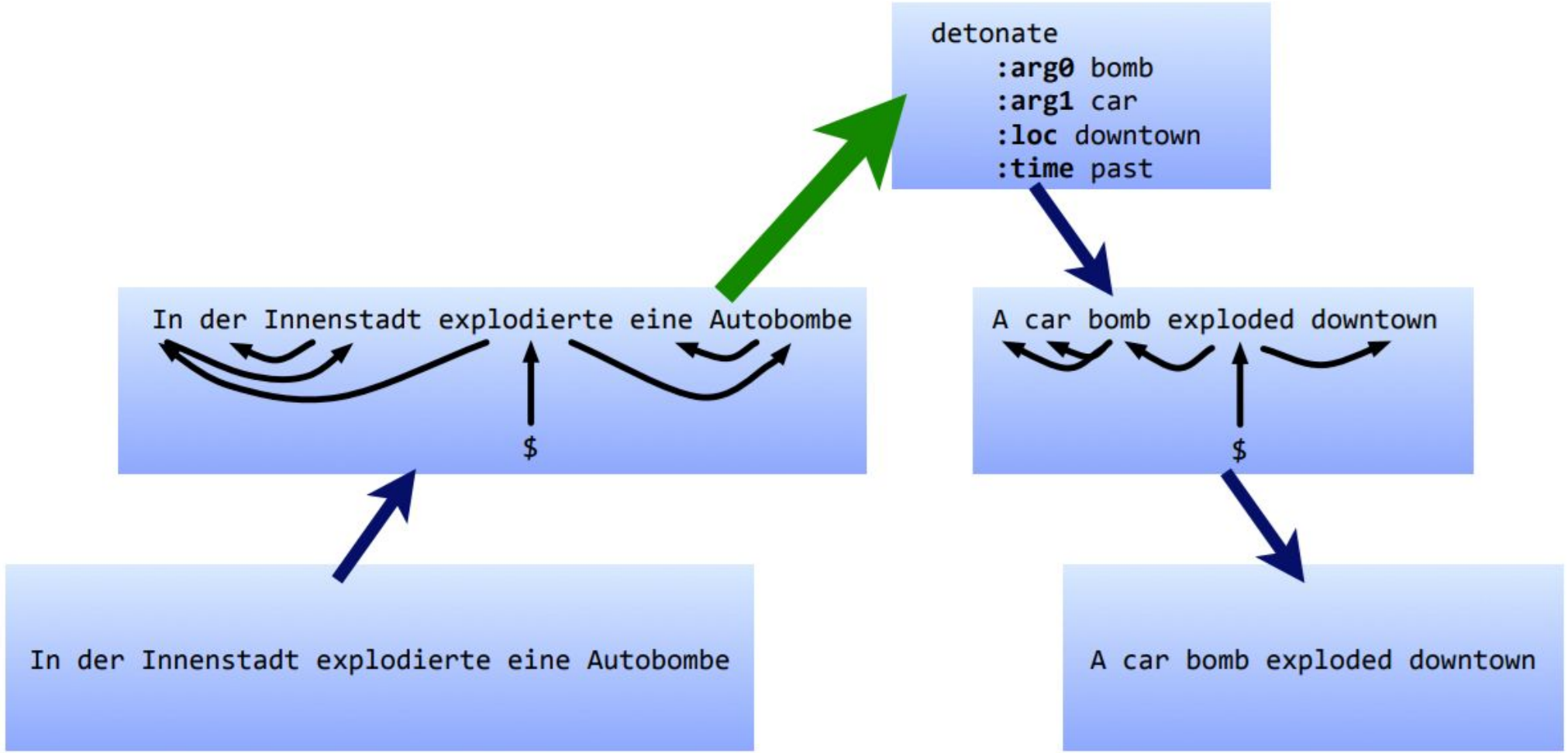
Levels of Transfer



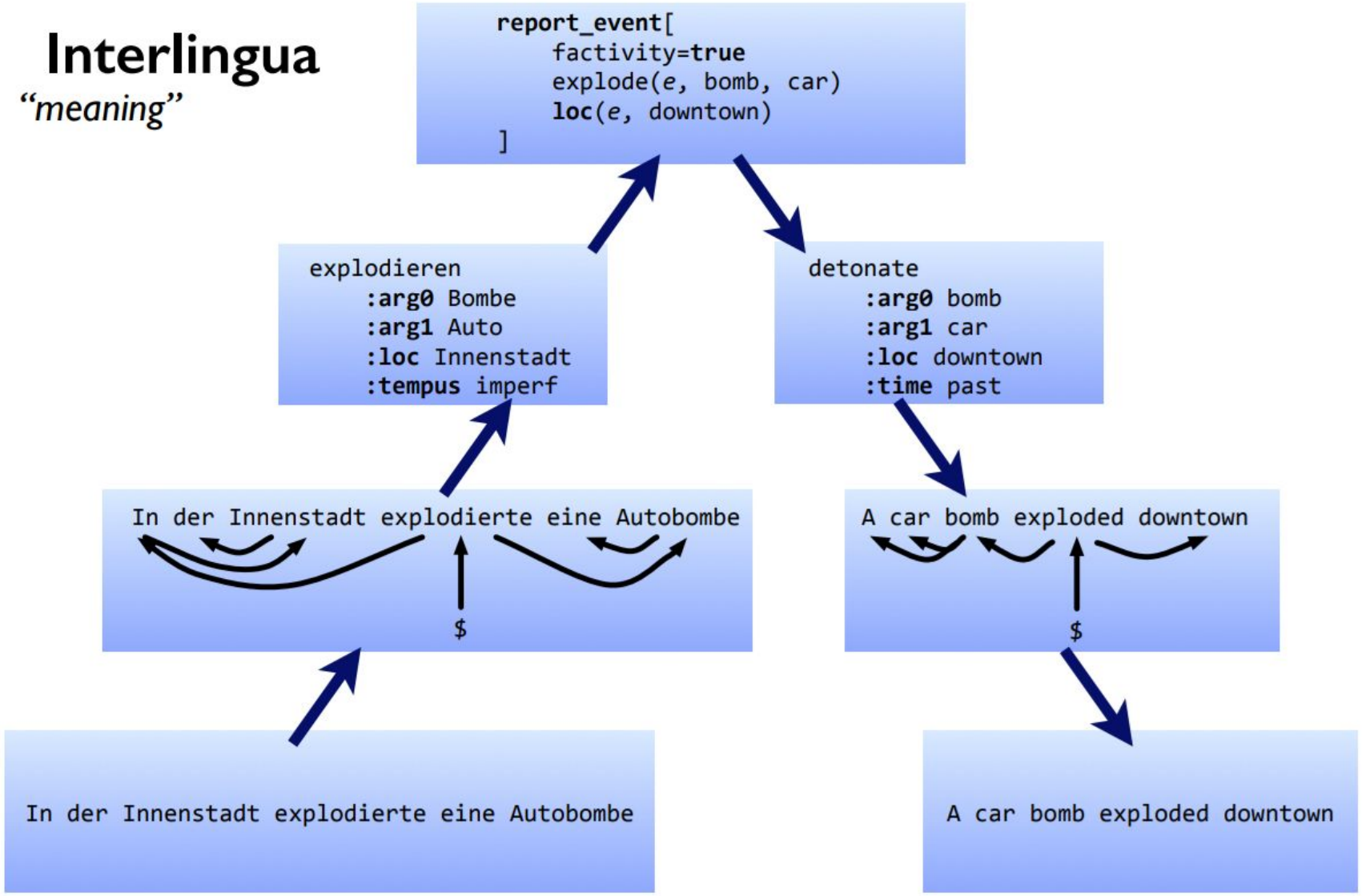
Levels of Transfer



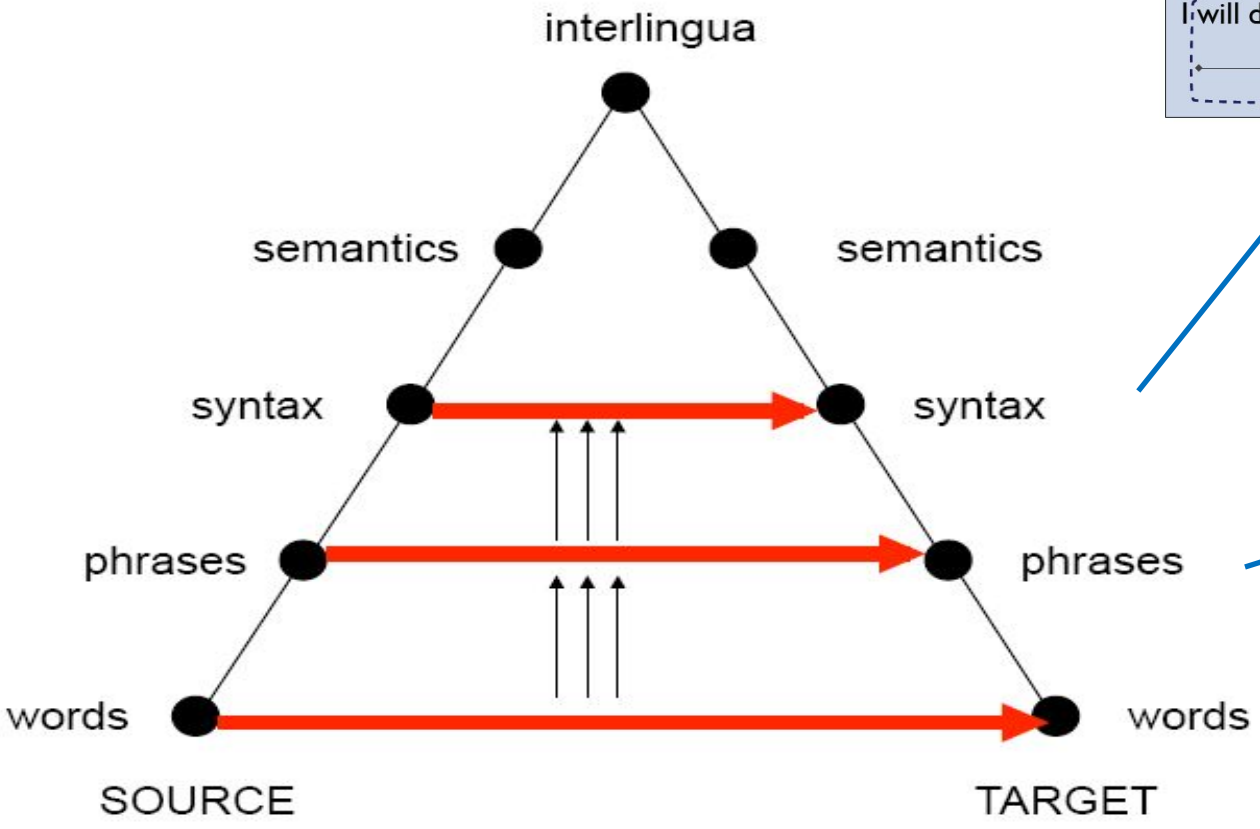
Levels of Transfer

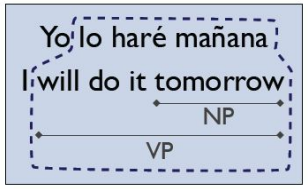


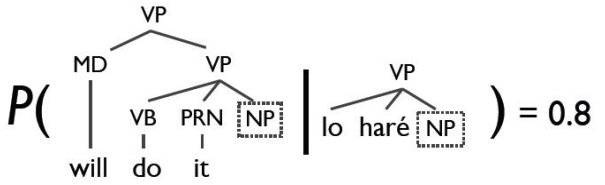
Levels of Transfer

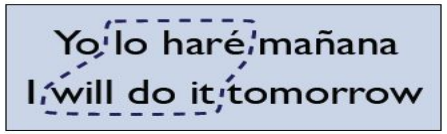


Levels of Transfer: The Vauquois triangle



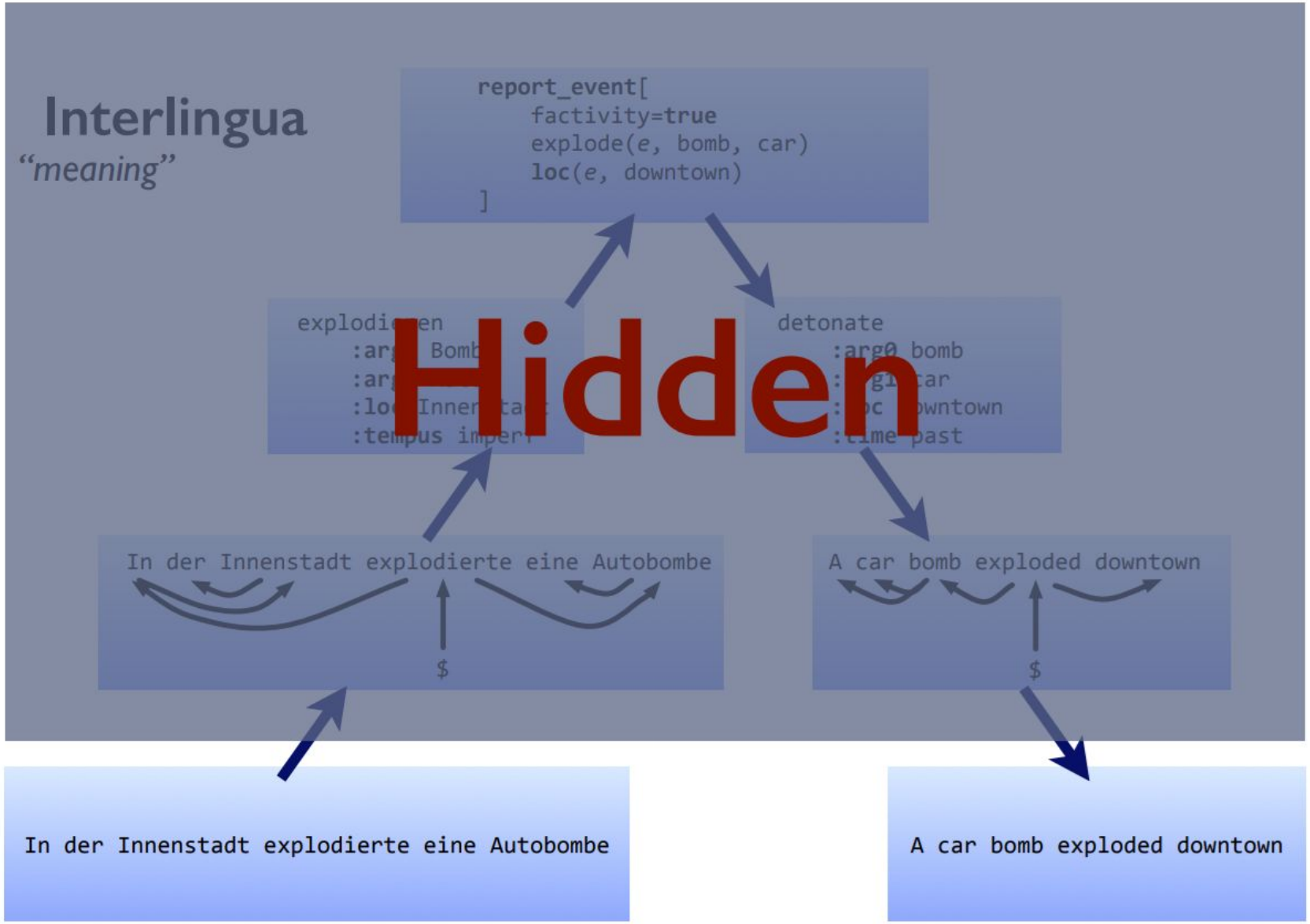


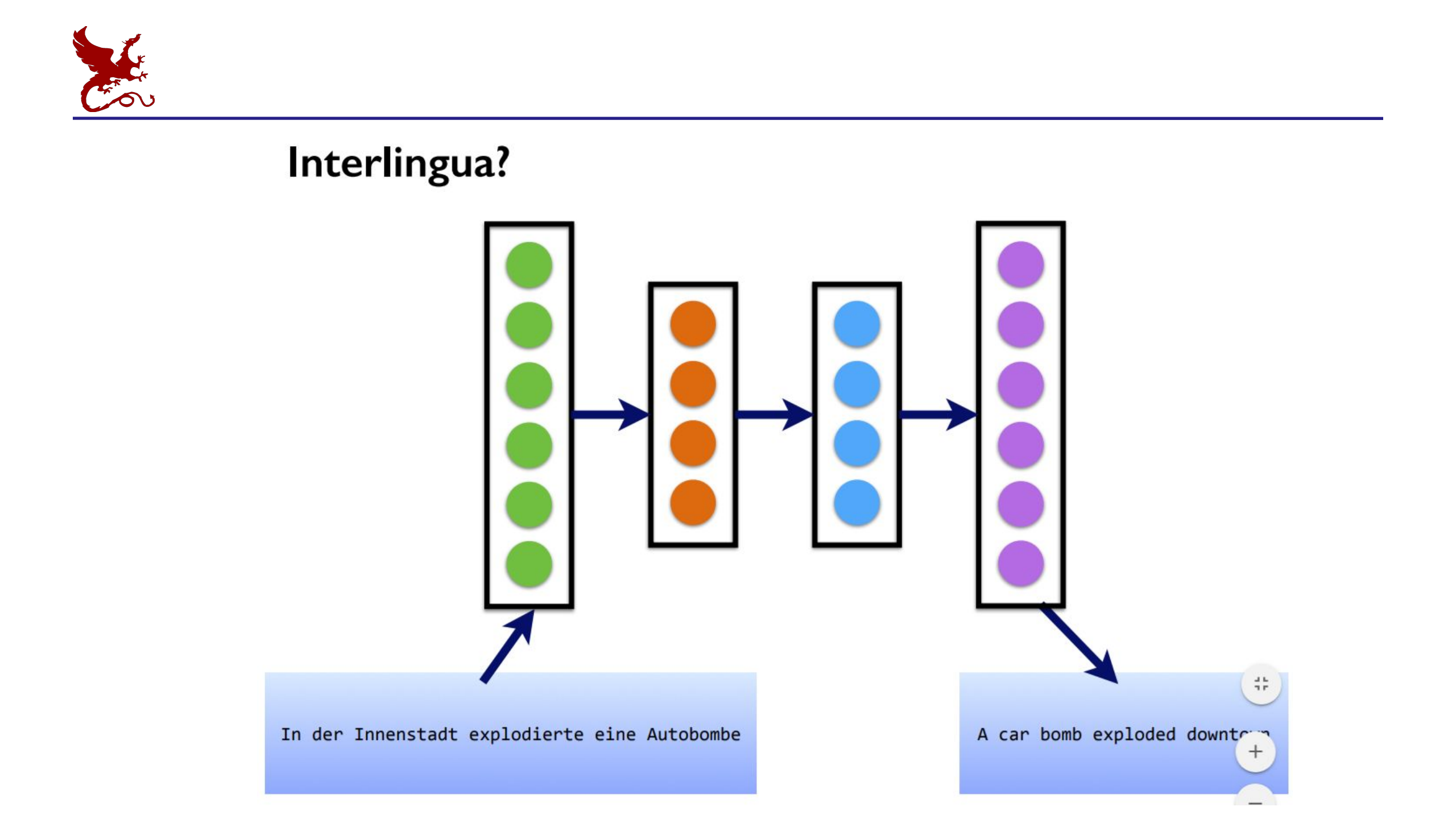




English (E)	*P*(E | lo haré)
will do it	0.8
will do so	0.2

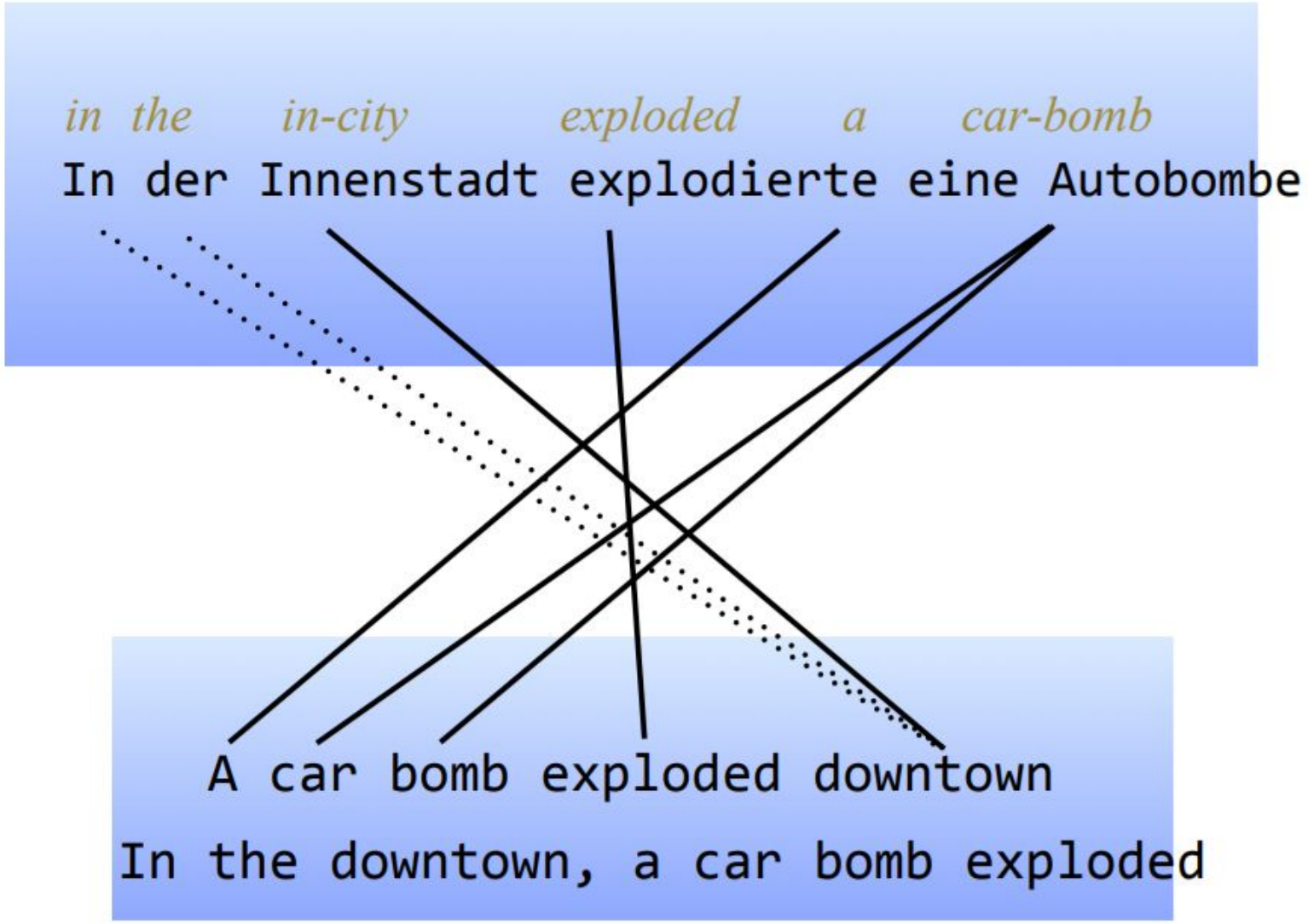
English (E)	*P*(E | mañana)
tomorrow	0.7
morning	0.3


Interlingua
"meaning"
report_event[
factivity=true
explode(e, bomb, car)
loc(e, downtown)
]
detonate
:arg0 bomb
:time past
Hidden
In der Innenstadt explodierte eine Autobombe
$
A car bomb exploded downtown
$
In der Innenstadt explodierte eine Autobombe
A car bomb exploded downtown


Interlingua?
In der Innenstadt explodierte eine Autobombe
A car bomb exploded

Ambiguities

- words
- morphology
- syntax
- semantics
- pragmatics



Machine Translation: Examples

Atlanta, preso il killer del palazzo di Giustizia

ATLANTA - La grande paura che per 26 ore ha attanagliato Atlanta è finita: Brian Nichols, l'uomo che aveva ucciso tre persone a palazzo di Giustizia e che ha poi ucciso un agente di dogana, s'è consegnato alla polizia, dopo avere cercato rifugio nell'alloggio di una donna in un complesso d'appartamenti alla periferia della città. Per tutto il giorno, il centro della città, sede della Coca Cola e dei Giochi 1996, cuore di una popolosa area metropolitana, era rimasto paralizzato.

Atlanta, taken the killer of the palace of Justice

ATLANTA - The great fear that for 26 hours has gripped Atlanta is ended: Brian Nichols, the man who had killed three persons to palace of Justice and that a customs agent has then killed, s' is delivered to the police, after to have tried shelter in the lodging of one woman in a complex of apartments to the periphery of the city. For all the day, the center of the city, center of the Coke Strains and of Giochi 1996, heart of one popolosa metropolitan area, was remained paralyzed.

Word-Level MT: Examples

la politique de la haine . *(Foreign Original)*
politics of hate . (Reference Translation)
the policy of the hatred . (IBM4+N-grams+Stack)

nous avons signé le protocole . *(Foreign Original)*
we did sign the memorandum of agreement . (Reference Translation)
we have signed the protocol . (IBM4+N-grams+Stack)

où était le plan solide ? *(Foreign Original)*
but where was the solid plan ? (Reference Translation)
where was the economic base ? (IBM4+N-grams+Stack)

Phrasal MT: Examples

Le président américain Barack Obama doit annoncer lundi de nouvelles mesures en faveur des constructeurs automobile. General motors et Chrysler avaient déjà bénéficié fin 2008 d'un prêt d'urgence cumulé de 17,4 milliards de dollars, et ont soumis en février au Trésor un plan de restructuration basé sur un total de 22 milliards de dollars d'aides publiques supplémentaires.

Interrogé sur la chaîne CBS dimanche, le président a toutefois clairement précisé que le gouvernement ne preterait pas d'argent sans de fortes contreparties. *"Il faudra faire des sacrifices à tous les niveaux"*, a-t-il prévenu. *"Tout le monde devra se réunir autour de la table et se mettre d'accord sur une restructuration en profondeur"*.

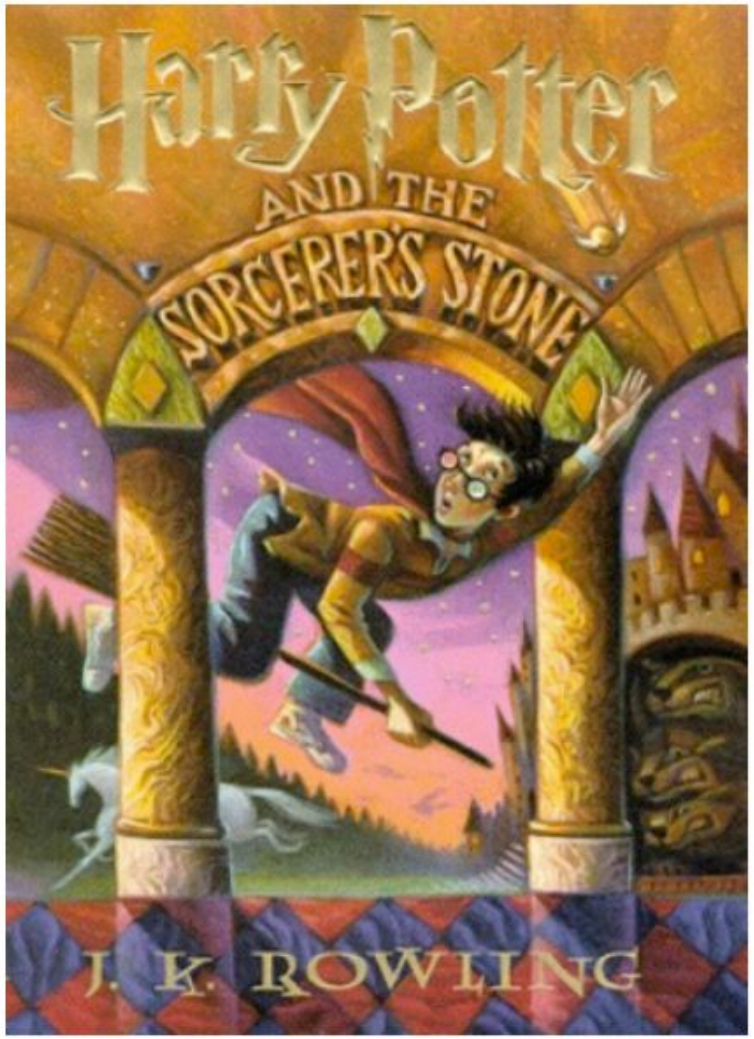
General Motors et Chrysler sont engagés dans des négociations avec le principal syndicat de l'automobile. Les constructeurs souhaitent diminuer leurs cotisations aux caisses de retraites, et accorder en échange des actions aux syndicats. Ils souhaiteraient également négocier des baisses des salaires.

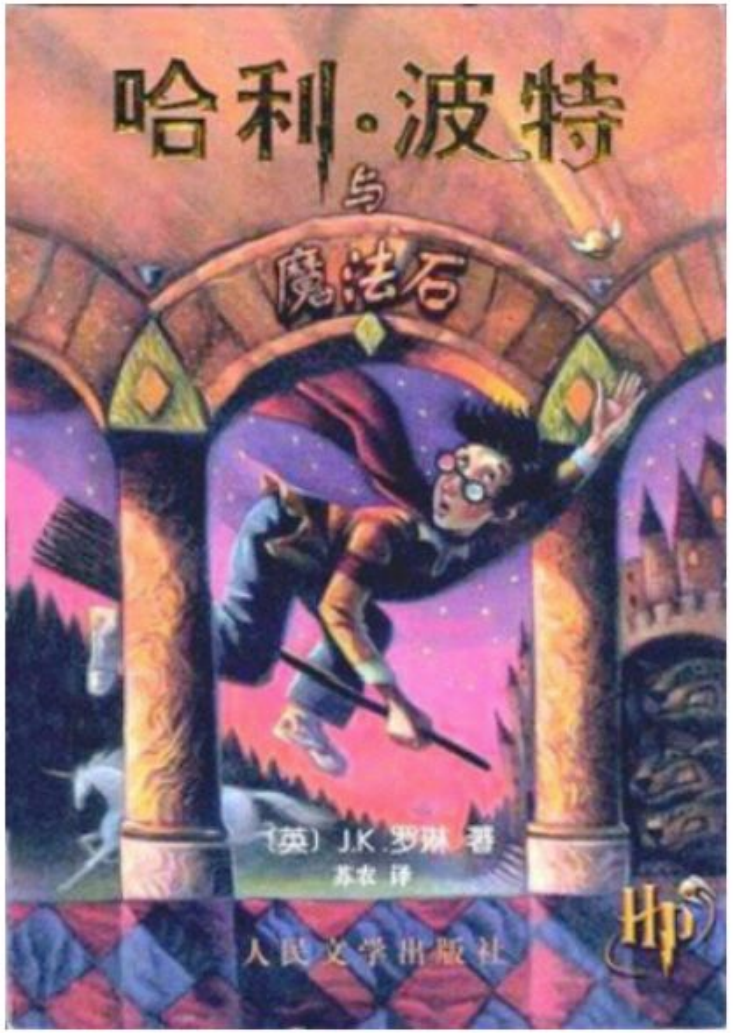
U.S. President Barack Obama to announce Monday new measures to help automakers. General Motors and Chrysler had already received late in 2008 a cumulative emergency loan of 17.4 billion dollars, and submitted to the Treasury in February in a restructuring plan based on a total of 22 billion dollars in additional aid .

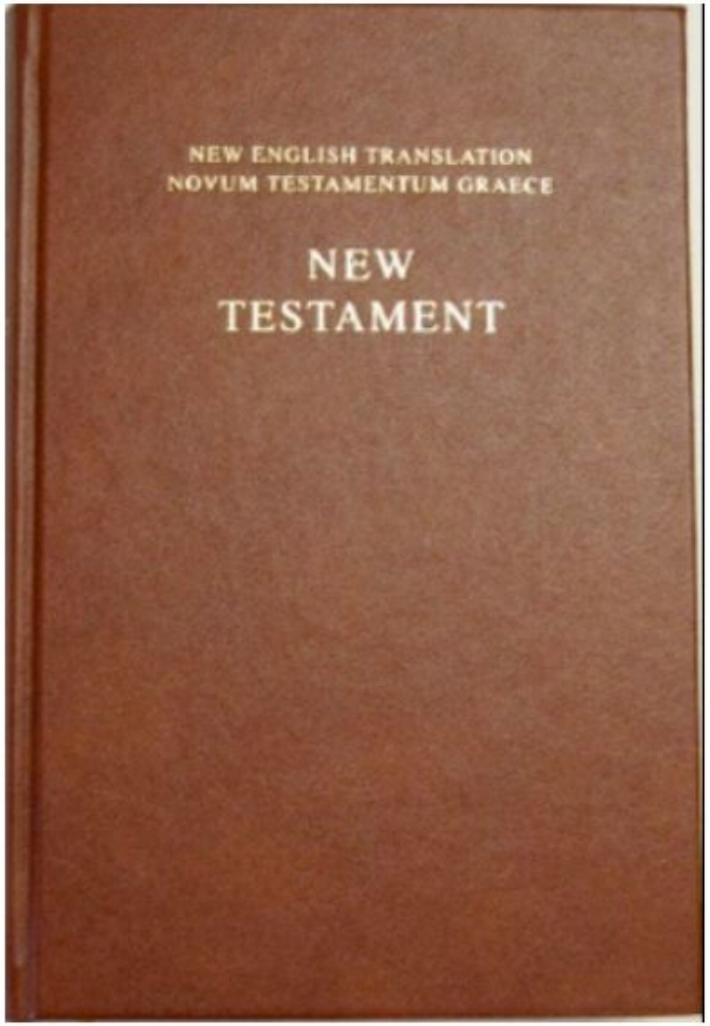
Interviewed on CBS Sunday, the president has clearly stated that the government does not lend money without strong counterparts. *"We must make sacrifices at all levels,"* he warned. *"Everyone should gather around the table and agree on a profound restructuring. "*

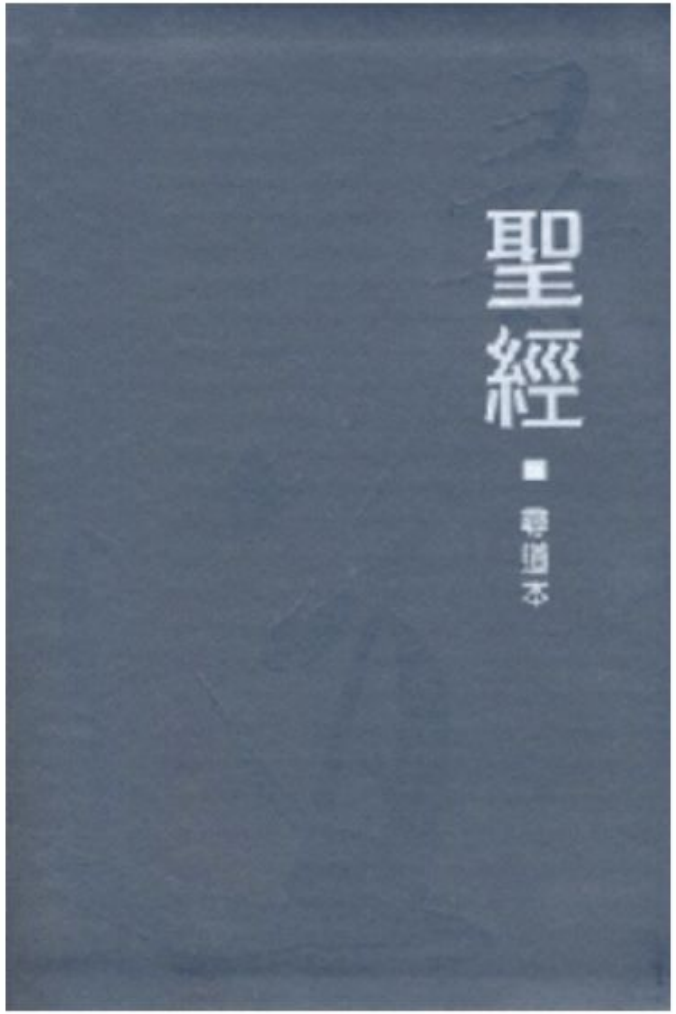
General Motors and Chrysler are engaged in negotiations with the major union of the car. Manufacturers wishing to reduce their contributions to pension funds, and give in exchange for the shares to trade unions. They would also negotiate lower wages.

Learning from Data


Harry Potter
AND THE
SORCERER'S STONE
J. K. ROWLING


哈利·波特
与
魔法石
(英) J.K.罗琳 著
苏农 译
人民文学出版社


NEW ENGLISH TRANSLATION
NOVUM TESTAMENTUM GRAECE
NEW
TESTAMENT


聖經
■
和合本

CLASSIC SOUPS

			Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50


www.un.org
http://www.un.org/english/
Google
We the peoples
Daily Briefing | Radio, TV, Photo | Documents, Maps | Publications, Stamps, Databases | UN Works | Search
Peace & Security | Economic & Social Development | Human Rights | Humanitarian Affairs | International Law
Welcome to the
United Nations
UN Millennium Development Goals
United Nations News Centre
About the United Nations
Main Bodies
Conferences & Events
Member States
General Assembly President
Secretary-General
Situation in Iraq
Mideast Roadmap
Renewing the UN
UN Action against Terrorism
Issues on the UN Agenda
Civil Society / Business
UN Webcast
CyberSchoolBus
8 September 2005 >>
Home | Recent Additions | Employment | UN Procurement | Comments | Q & A | UN System Sites | Index
عربي | 中文 | English | Français | Русский | Español
Copyright, United Nations, 2000-2005 | Use of UNBC Logo | Terms of Use | Privacy Notice | Help
[Text version]
Live and On-Demand Webcasts, 24 Hours a Day: Click on UN Webcas


联合国主页
http://www.un.org/chinese/
Google
我们人民
每日简报 | 多媒体 | 文件与地图 | 出版物 邮票 数据库 | 服务全球 | 网址搜索
和平与安全 | 经济与社会发展 | 人权 | 人道主义事务 | 国际法
欢迎来到
联合国
联合国
千年发展目标
联合国新闻
联合国概况
联合国主要机关
会议与活动
联合国会员国
联合国大会主席
联合国秘书长
伊拉克局势
中东路线图
更新联合国
反恐怖主义
联合国日常议题
民间团体/商业
联合国网络直播
空中校车
联大第60届会议一般性辩论
新增内容 | 工作机会 | 联合国采购 | 建议 | 问题与解答 | 其他网址 | 网址索引
عربي | 中文 | English | Français | Русский | Español
联合国2000-2005年版权|联合国60周年徽标使用准则|使用条件|隐私通告|帮助
[纯文字版]

http://opus.nlpl.eu

OPUS ... the open parallel corpus

OPUS is a growing collection of translated texts from the web. In the OPUS project we try to convert and align free online data, to add linguistic annotation, and to provide the community with a publicly available parallel corpus. OPUS is based on open source products and the corpus is also delivered as an open content package. We used several tools to compile the current collection. All pre-processing is done automatically. No manual corrections have been carried out.

The OPUS collection is growing! Check this page from time to time to see new data arriving ... Contributions are very welcome! Please contact <jorg.tiedemann@helsinki.fi >

Search & download resources: -- select -- | -- select -- | all

Latest News

- 2018-02-15: New corpora: ParaCrawl, XhosaNavy
- 2017-11-06: New version: OpenSubtitles2018
- 2017-11-01: New server location: http://opus.nlpl.eu
- 2016-01-08: New version: OpenSubtitles2016
- 2015-10-15: New versions of TED2013, NCv9
- 2014-10-24: New: JRC-Acquis
- 2014-10-20: NCv9, TED talks, DGT, WMT
- 2014-08-21: New: Ubuntu, GNOME
- 2014-07-30: New: Translated Books
- 2014-07-27: New: DOGC, Tanzil
- 2014-05-07: Parallel coref corpus ParCor

Search & Browse

- OPUS multilingual search interface
- Europarl v7 search interface
- Europarl v3 search interface
- OpenSubtitles 2016 search interface
- EUconst search interface
- Word Alignment Database (old DB)

Tools & Info

- OPUS Wiki
- OPUS API by Yonathan Koren
- Uplug at bitbucket

Some Projects using OPUS

- Let'sMT! - On-line SMT toolkit
- CASMACAT - Computer-Aided Translation

Sub-corpora (downloads & infos):

- Books - A collection of translated literature (Books.tar.gz - 535 MB)
- DGT - A collection of EU Translation Memories provided by the JRC
- DOGC - Documents from the Catalan Goverment (DOGC.tar.gz - 2.8 GB)
- ECB - European Central Bank corpus (ECB.tar.gz - 3.0 GB)
- EMEA - European Medicines Agency documents (EMEA.tar.gz - 13.0 GB)
- The EU bookshop corpus (EUbookshop.tar.gz - 42 GB)
- EUconst - The European constitution (EUconst.tar.gz - 82` MB)
- EUROPARL v7 - European Parliament Proceedings (Europarl.tar.gz - 2(GB)
- GNOME - GNOME localization files (GNOME.tar.gz - 9 GB)
- Global Voices - News stories in various languages (GlobalVoices.tar.gz - 1.2 GB)
- The Croatian - English WaC corpus (hrenWaC.tar.gz - 59 MB)
- JRC-Acquis- legislative EU texts (JRC-Acquis.tar.gz - 11 GB)

Learning from Data:
The Noisy Channel

$$\hat{\boldsymbol{e}} = \arg\max_{\boldsymbol{e}} p_{\boldsymbol{\varphi}}(\boldsymbol{e}) \times p_{\boldsymbol{\theta}}(\boldsymbol{f} \mid \boldsymbol{e})$$ Noisy channel

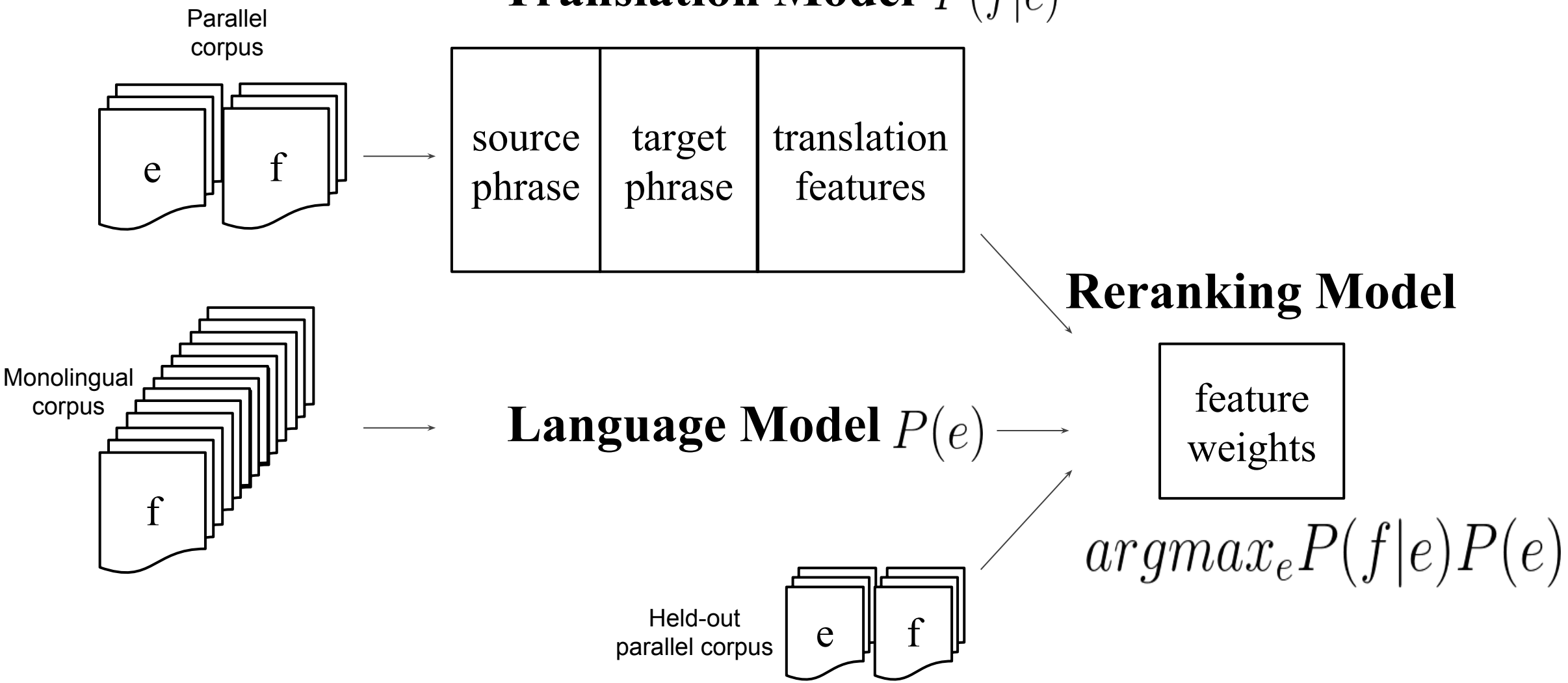
$$\hat{\boldsymbol{e}} = \arg\max_{\boldsymbol{e}} p_{\boldsymbol{\lambda}}(\boldsymbol{e} \mid \boldsymbol{f})$$ Direct

$$\hat{e} = \arg\max_{e} p_{\varphi}(e) \times p_{\theta}(f \mid e) \qquad \text{Noisy channel}$$

$$\hat{e} = \arg\max_{e} p_{\lambda}(e \mid f) \qquad \text{Direct}$$

- There is a lot more monolingual data in the world than translated data
- Easy to get about 1 trillion words of English by crawling the web
- With some work, you can get 1 billion translated words of English-French
  - What about English-German?
  - What about Japanese-Turkish?

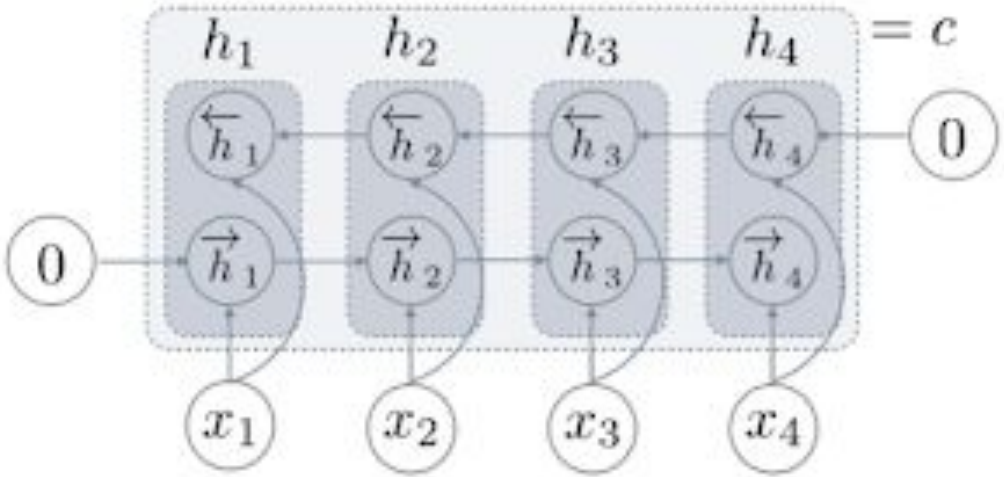
Phrase-Based MT



Neural MT: Conditional Language Modeling

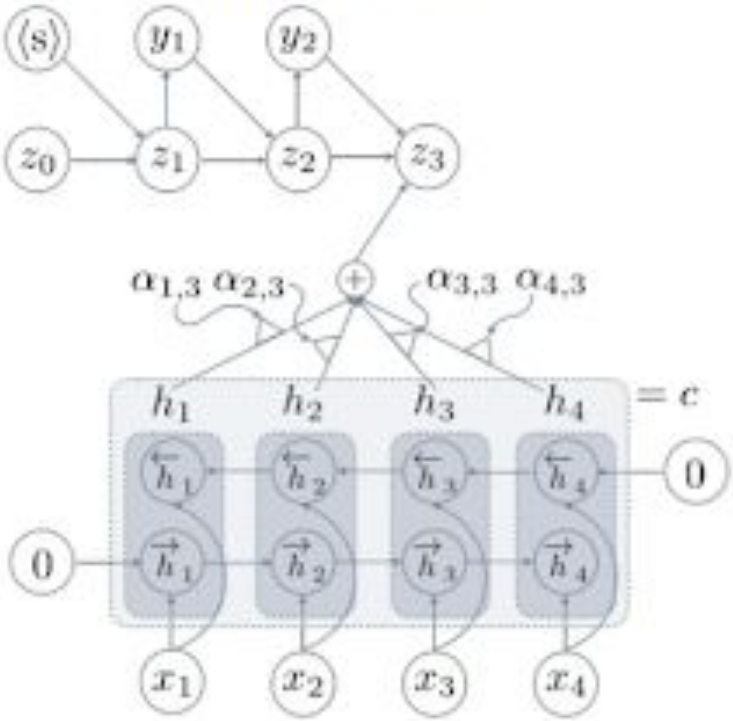
Encoder

- Project a source sentence into a set of continuous vectors



Decoder+Attention

- Decode a target sentence from a set of "source" continuous vectors



Slide credit: Kyunghyun Cho

Research Problems

- How can we formalize the process of learning to translate from examples?
- How can we formalize the process of finding translations for new inputs?
- If our model produces many outputs, how do we find the best one?
- If we have a gold standard translation, how can we tell if our output is good or bad?

MT Evaluation is Hard

- Language variability: there is no single correct translation
- Human evaluation is subjective
- How good is good enough? Depends on the application of MT (publication, reading, …)
- Is system A better than system B?
- MT Evaluation is a research topic on its own.
  - How do we do the evaluation?
  - How do we measure whether an evaluation method is good?

Human Evaluation

- Adequacy and Fluency
  - Usually on a Likert scale (1 “not adequate at all” to 5 “completely adequate”)
- Ranking of the outputs of different systems at the system level

Human Evaluation

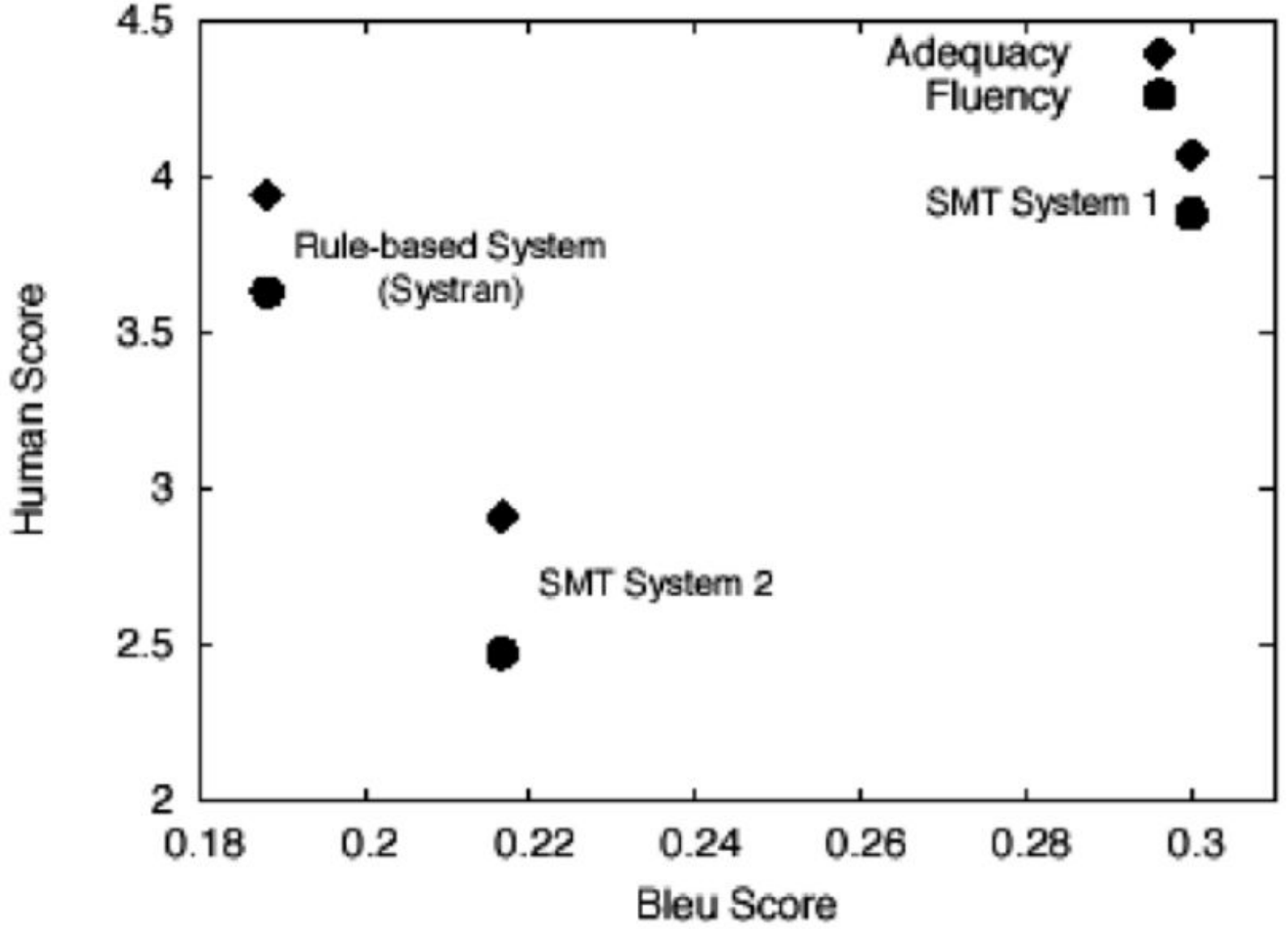
- Adequacy and Fluency
  - Usually on a Likert scale (1 “not adequate at all” to 5 “completely adequate”)
- Ranking of the outputs of different systems at the system level

- Post editing effort: how much effort does it take for a translator (or even monolingual) to “fix” the MT output so it is “good”
- Task-based evaluation: was the performance of the MT system sufficient to perform a task.

Automatic Evaluation

- The BLEU score proposed by IBM (Papineni et al., 2002)
  - Exact matches of n-grams
  - Match against a set of reference translations for greater discrimination between good and bad translations
  - Account for adequacy by looking at word precision
  - Account for fluency by calculating n-gram precisions for n=1,2,3,4
  - No recall (because difficult with multiple references)
  - To compensate for recall: “brevity penalty”. Translates that are too short are penalized
  - Final score is the geometric average of the n-gram precisions, times the brevity penalty
  - Calculate the aggregate score over a large test set

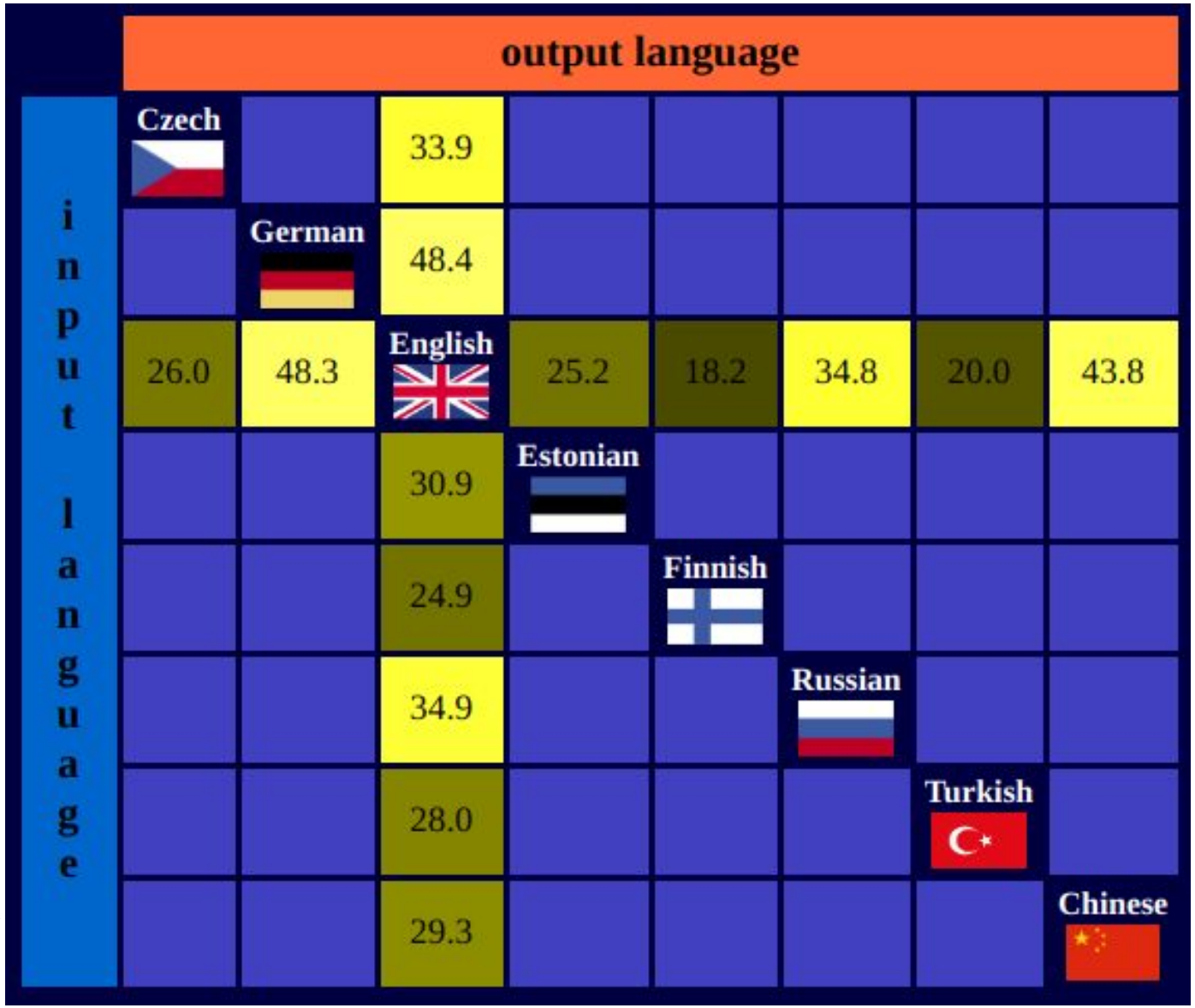
BLEU vs. Human Scores



BLEU Scores

- More reference human translations results in better and more accurate scores
- General interpretability of scale
  - Scores over 30 (single reference) are generally understandable
  - Scores over 50 (single reference) are generally good and fluent

WMT 2018



http://www.statmt.org/wmt18/

Systems Overview

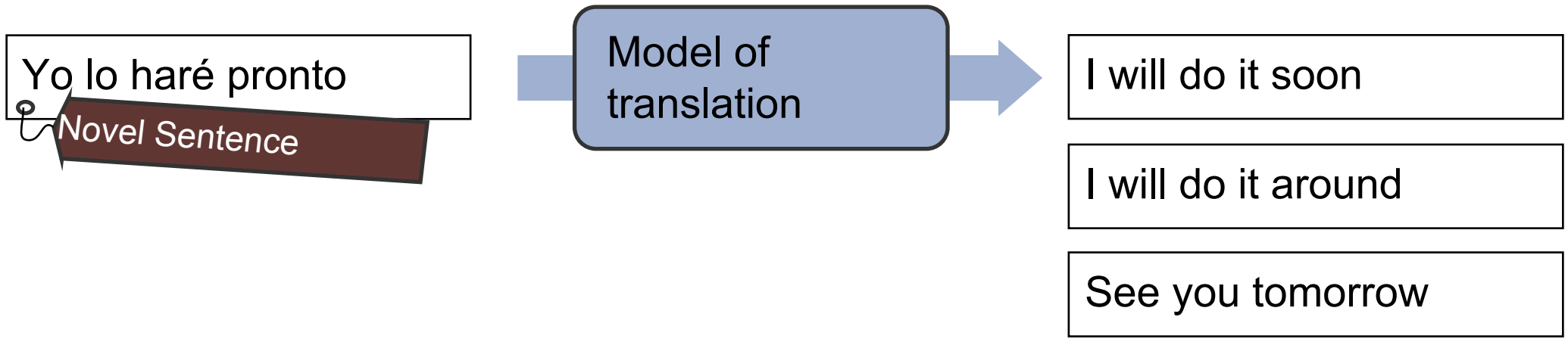
Corpus-Based MT

Modeling correspondences between languages

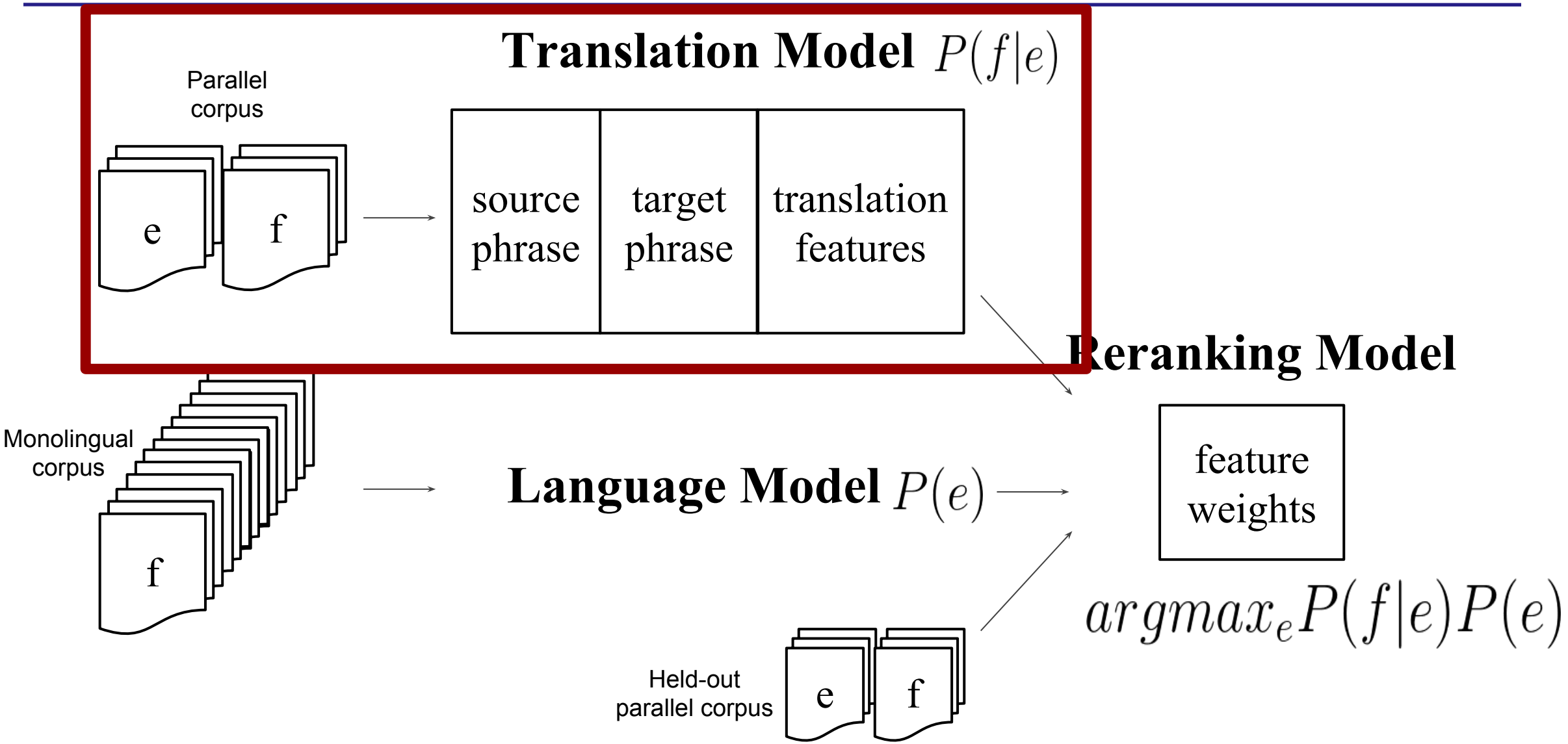
Sentence-aligned parallel corpus:

Yo lo haré mañana I will do it tomorrow	Hasta pronto See you soon	Hasta pronto See you around

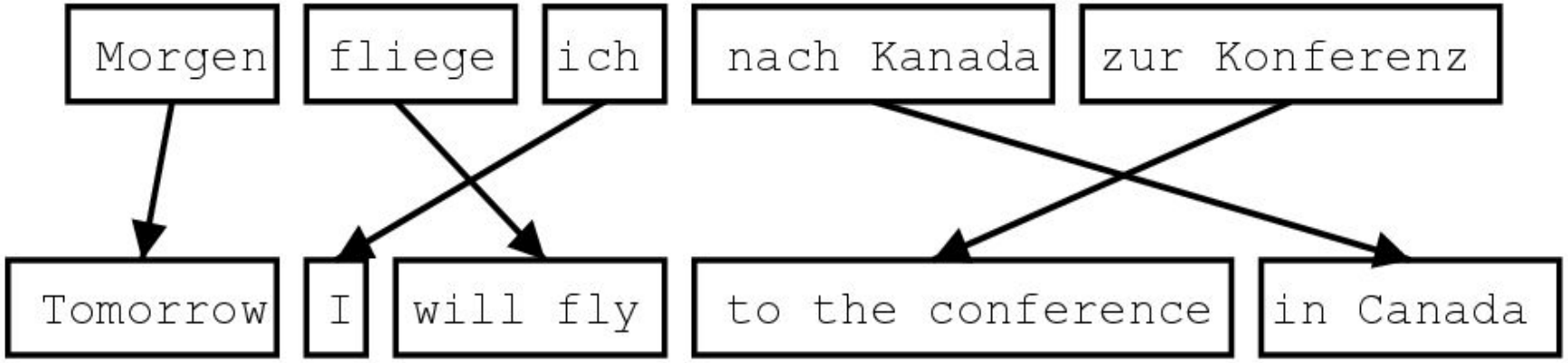
Machine translation system:

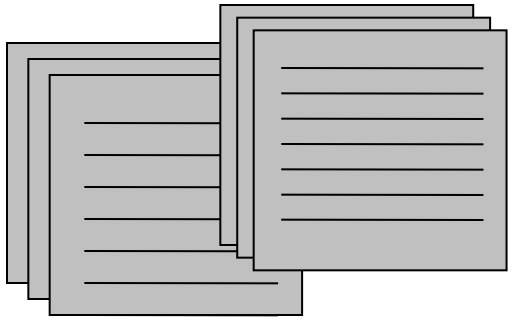


Phrase-Based MT

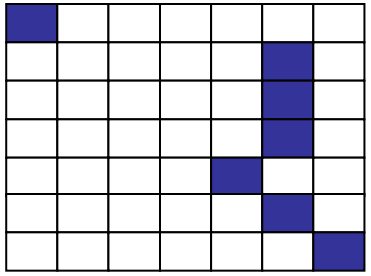


Phrase-Based System Overview





Sentence-aligned corpus



Word alignments

```
cat ||| chat ||| 0.9
the cat ||| le chat ||| 0.8
dog ||| chien ||| 0.8
house ||| maison ||| 0.6
my house ||| ma maison ||| 0.9
language ||| langue ||| 0.9
…
```

Phrase table (translation model)

Many slides and examples from Philipp Koehn or John DeNero

Word Alignment

Lexical Translation

- How do we translate a word? Look it up in the dictionary

  Haus : house, home, shell, household

- Multiple translations
  - Different word senses, different registers, different inflections (?)
  - *house, home* are common
- *shell* is specialized (the Haus of a snail is a shell)

How common is each?

Translation	Count
house	5000
home	2000
shell	100
household	80

MLE

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

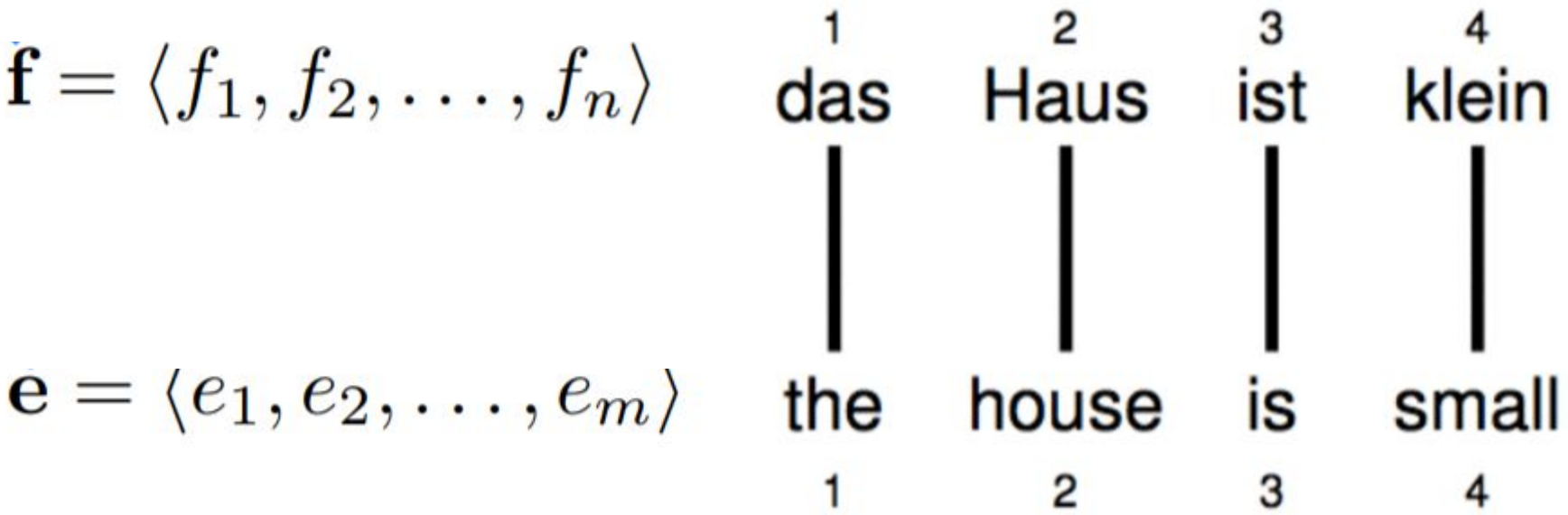
Lexical Translation

- Goal: a model $p(\mathbf{e} \mid \mathbf{f}, m)$
- where **e** and **f** are complete English and Foreign sentences

$$\mathbf{e} = \langle e_1, e_2, \ldots, e_m \rangle \qquad \mathbf{f} = \langle f_1, f_2, \ldots, f_n \rangle$$

The Alignment Function

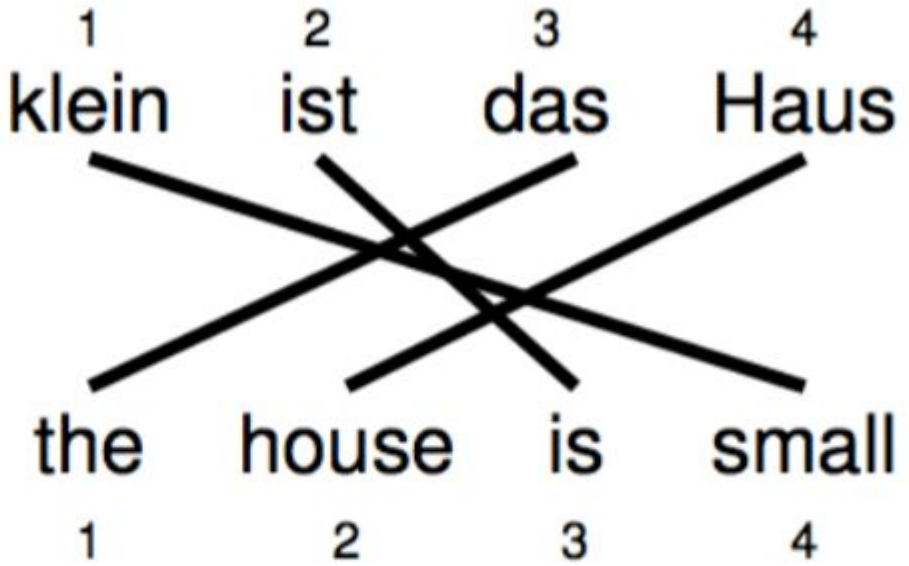
- Alignments can be visualized in by drawing links between two sentences, and they are represented as vectors of positions:



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

Reordering

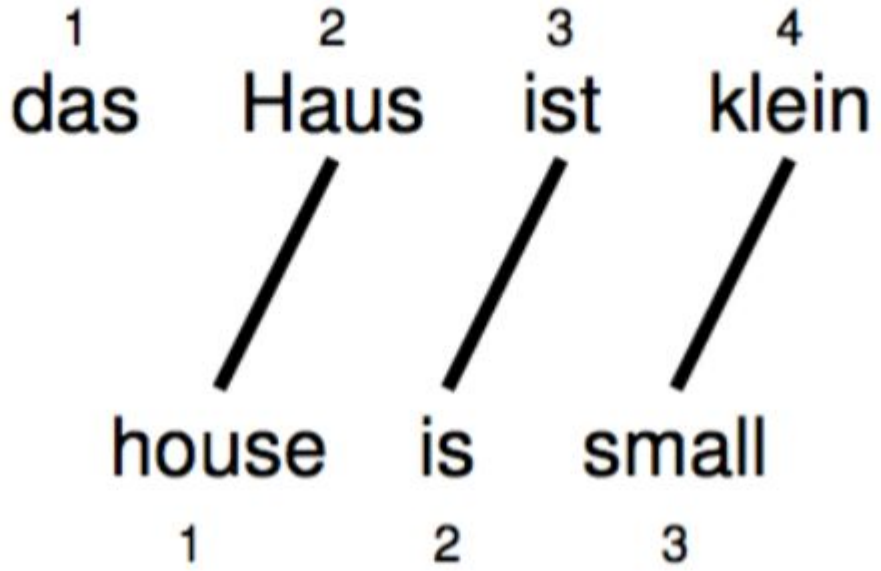
- Words may be reordered during translation.



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

Word Dropping

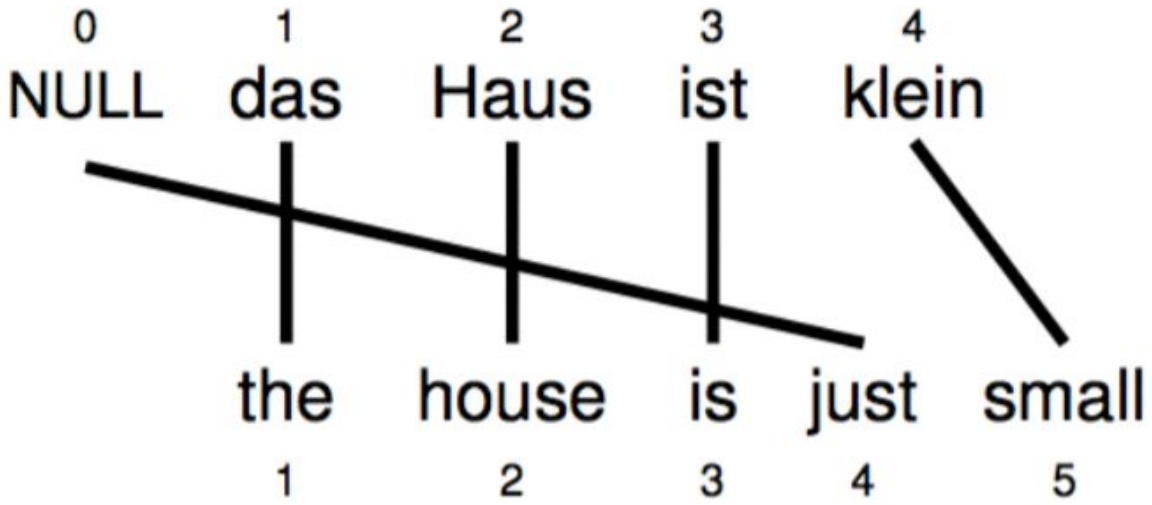
- A source word may not be translated at all



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

Word Insertion

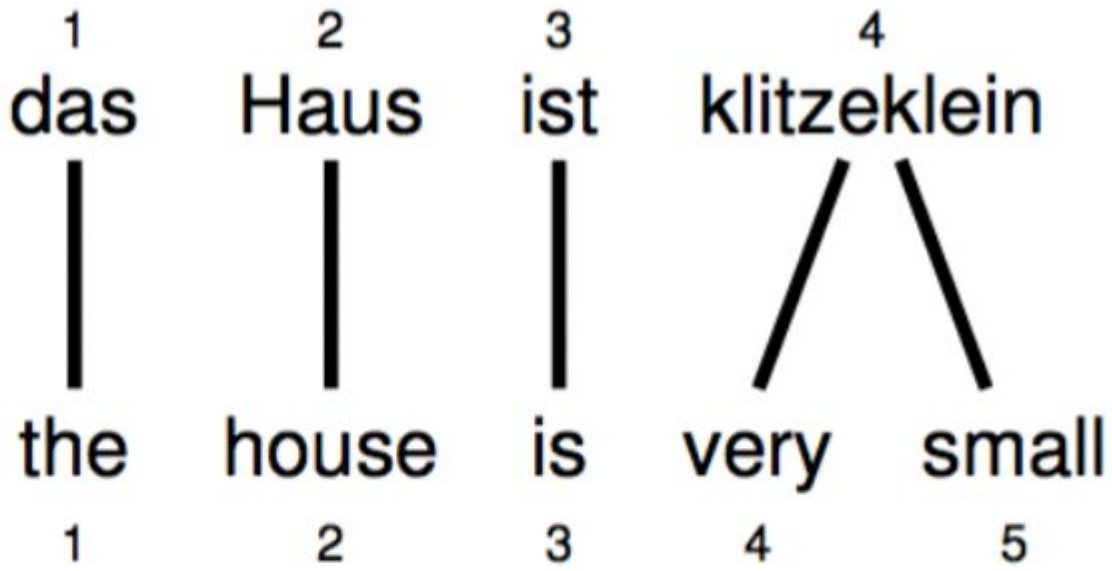
- Words may be inserted during translation
  - English *just* does not have an equivalent
  - But it must be explained - we typically assume every source sentence contains a NULL token



$$\mathbf{a} = (1, 2, 3, 0, 4)^\top$$

One-to-many Translation

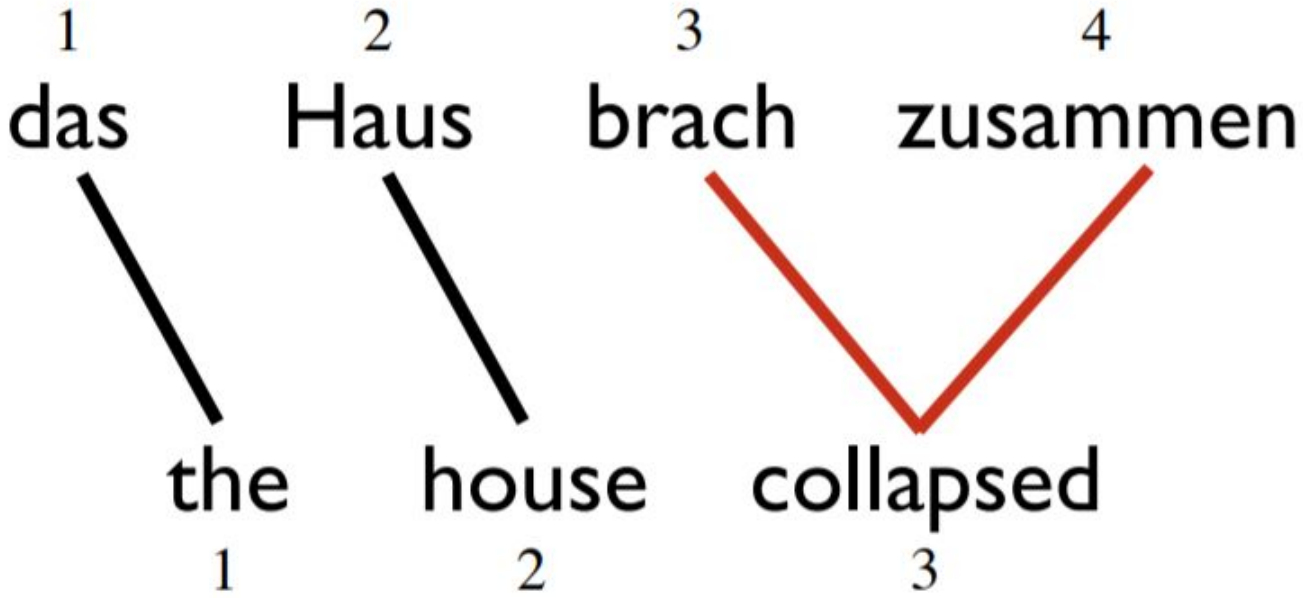
- A source word may translate into more than one target word



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

Many-to-one Translation

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



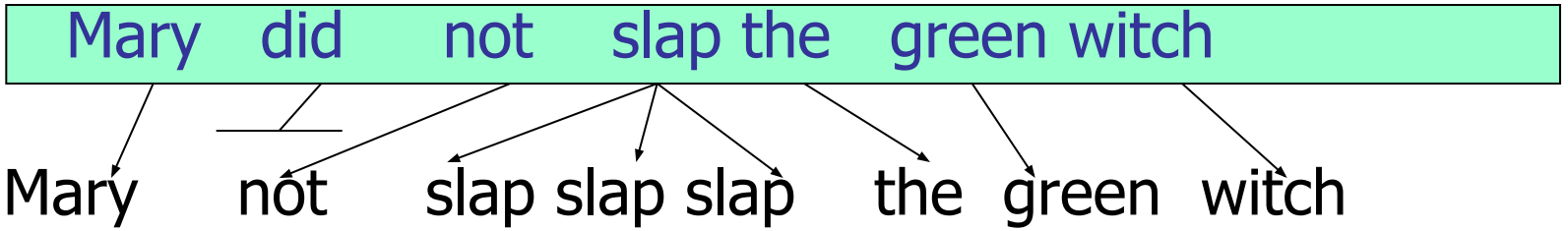
$\mathbf{a} = ???$ $\mathbf{a} = (1, 2, (3, 4)^\top)^\top$?

Generative Story

$$p(\mathbf{e} \mid \mathbf{f}, m) \; ?$$

Mary did not slap the green witch

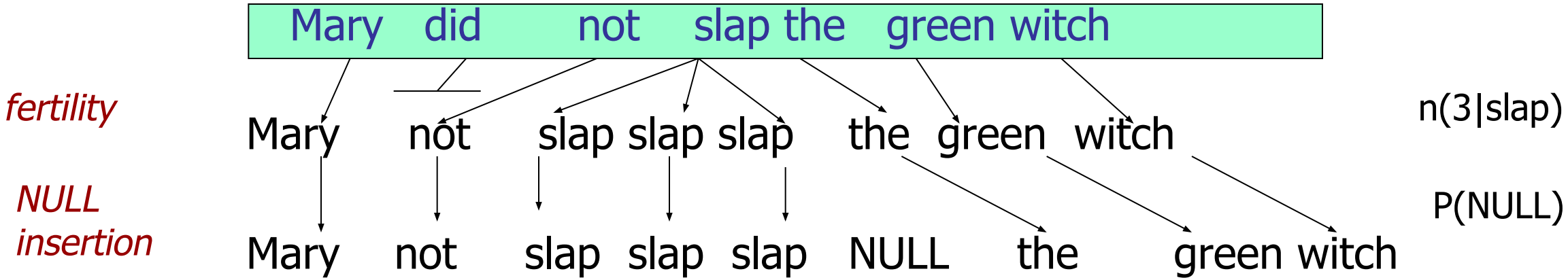
Generative Story



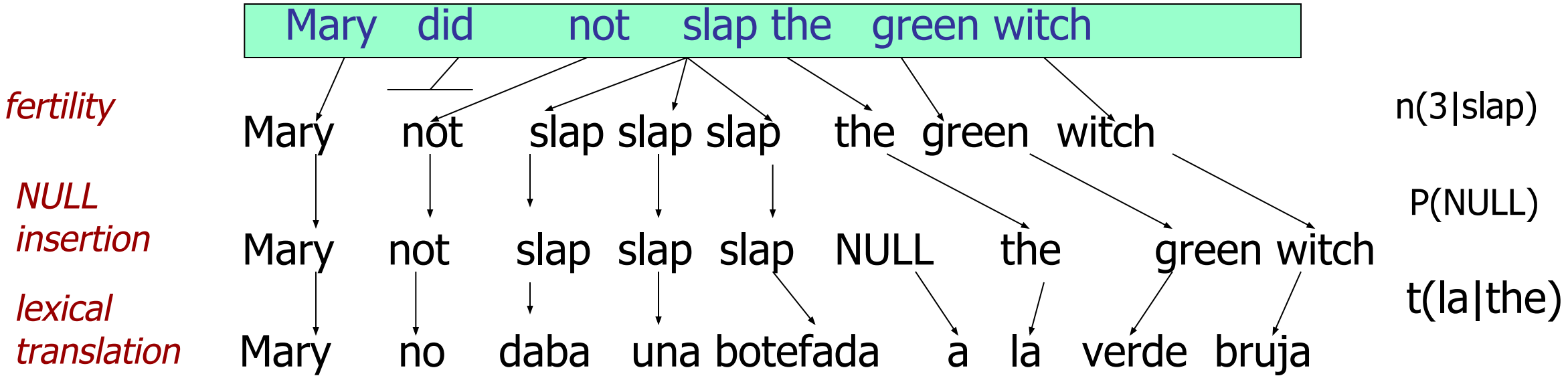
fertility

n(3|slap)

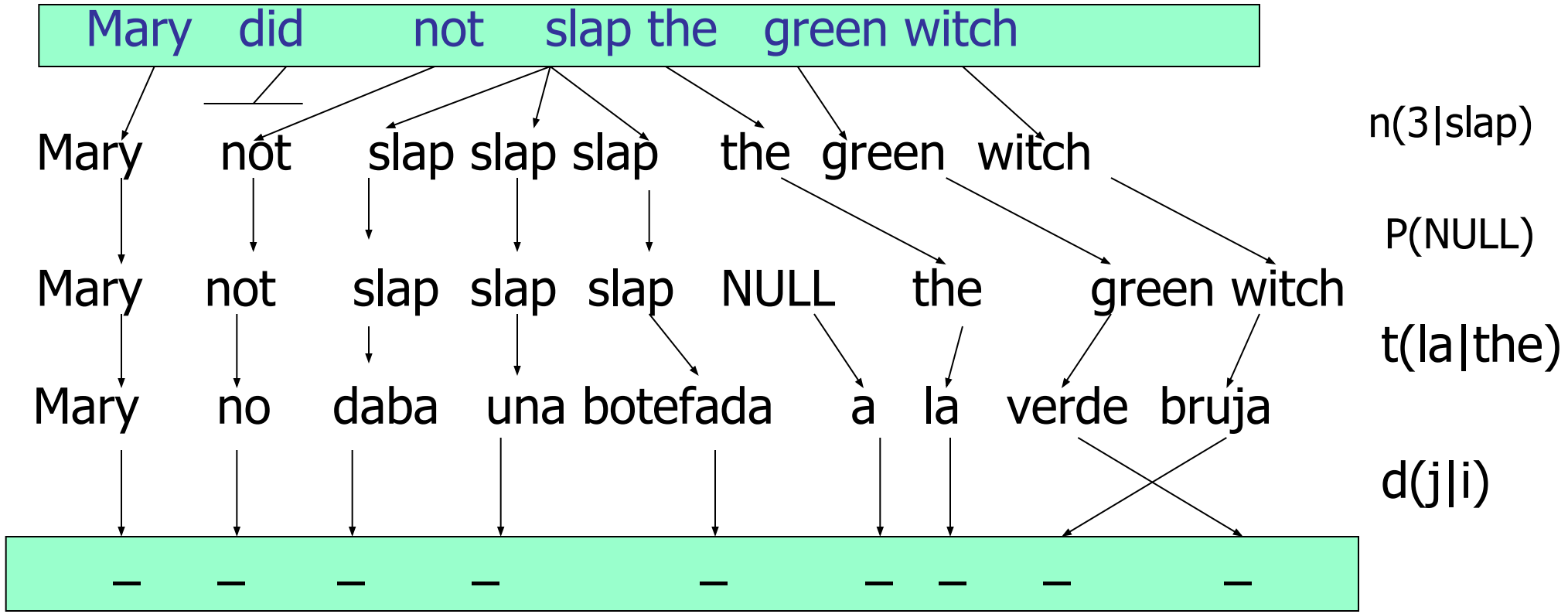
Generative Story



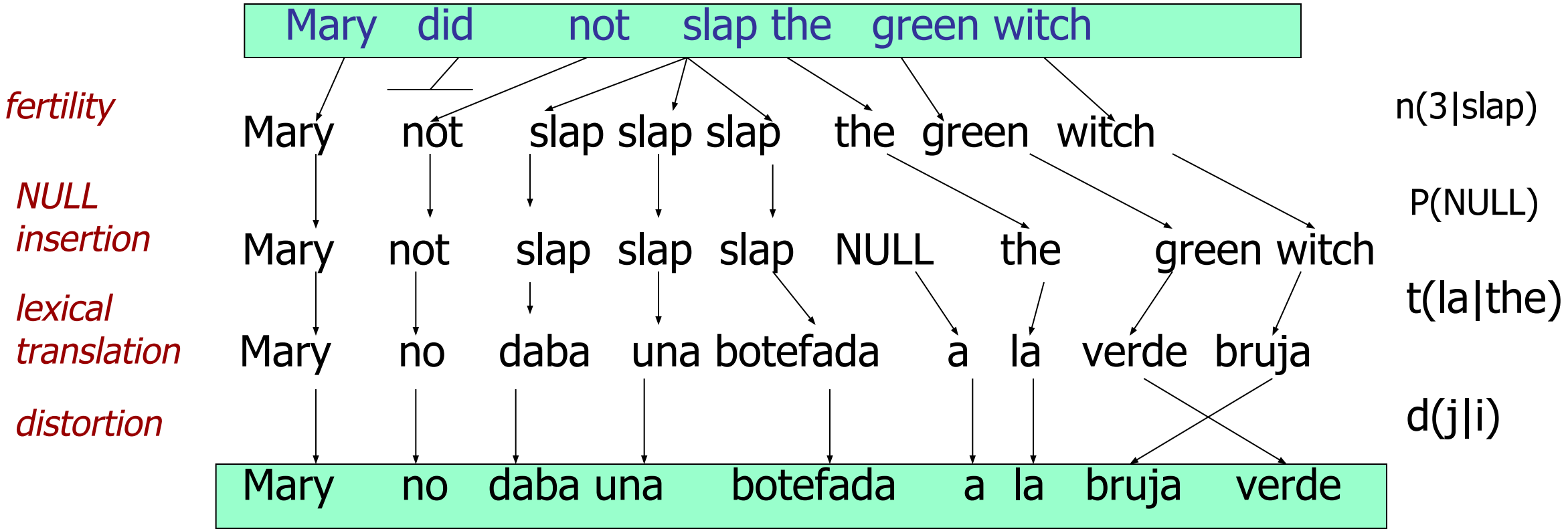
Generative Story



Generative Story



The IBM Models (Brown et al. 93)

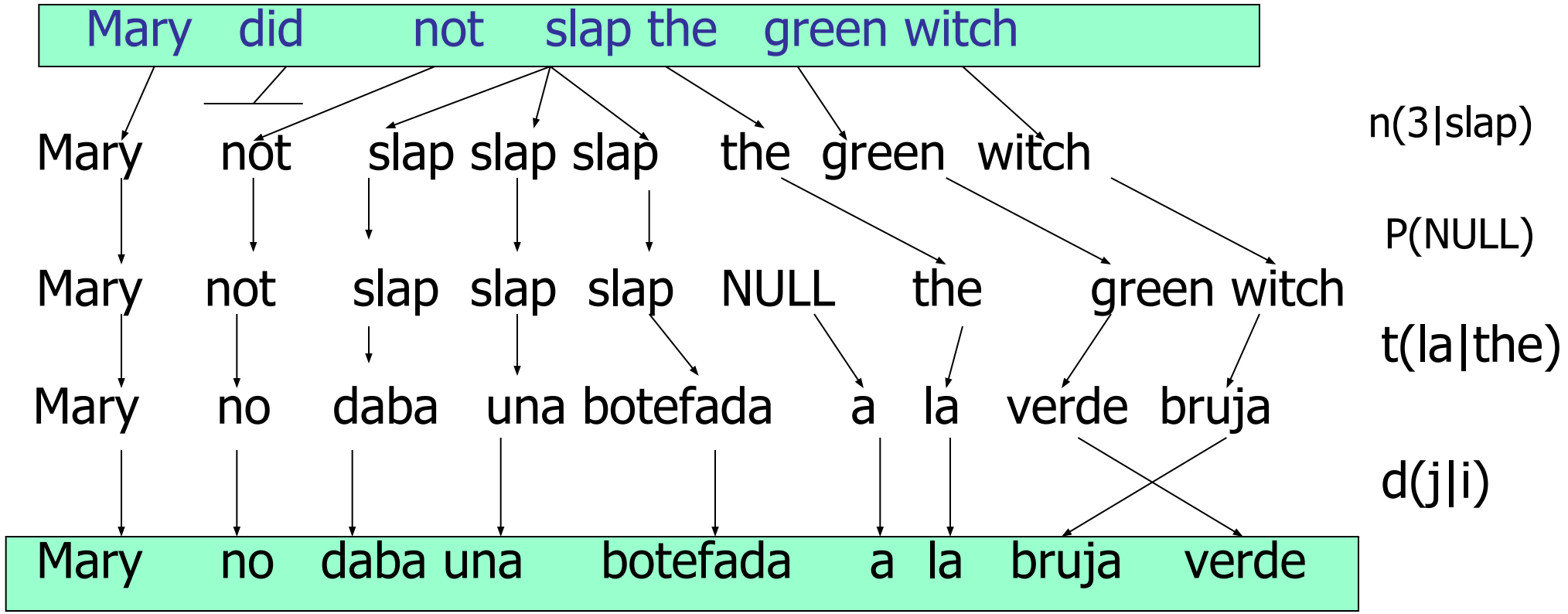


[from Al-Onaizan and Knight, 1998]

Alignment Models

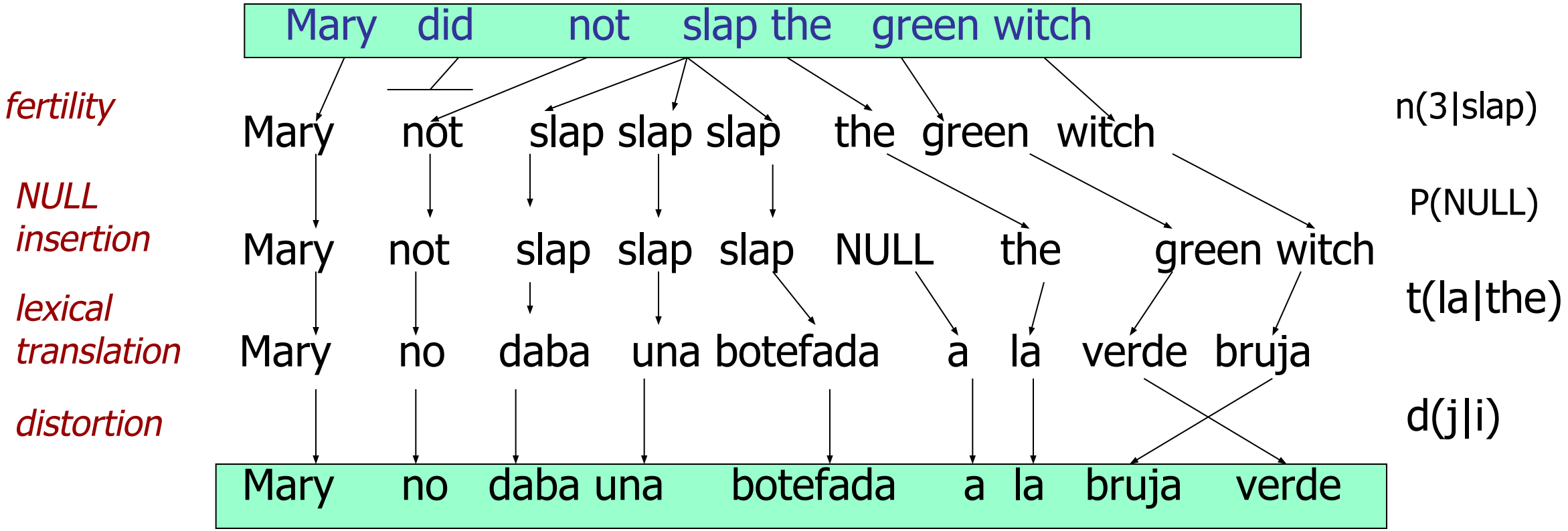
- IBM Model 1: lexical translation
- IBM Model 2: alignment model, global monotonicity
- HMM model: local monotonicity
- *fastalign:* efficient reparametrization of Model 2
- IBM Model 3: fertility
- IBM Model 4: relative alignment model
- IBM Model 5: deficiency
- …

P(e,a|f)



$$P(e, \text{alignment}|f) = \prod p_f \prod p_t \prod p_d$$

P(e|f)

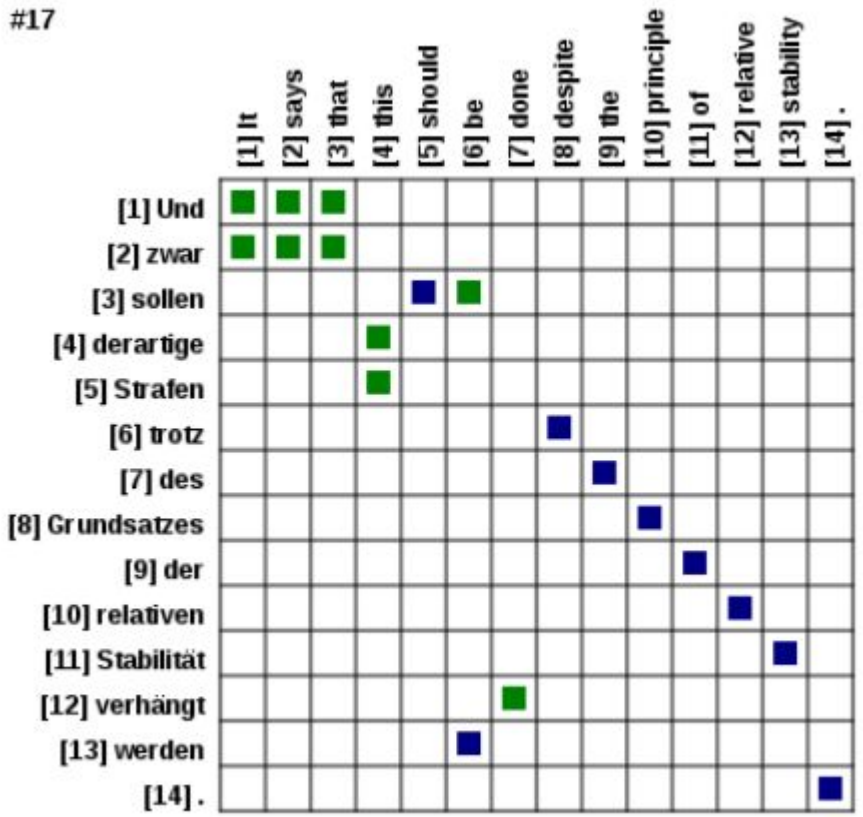


$$P(e|f) = \sum_{all\_possible\_alignments} \prod p_f \prod p_t \prod p_d$$

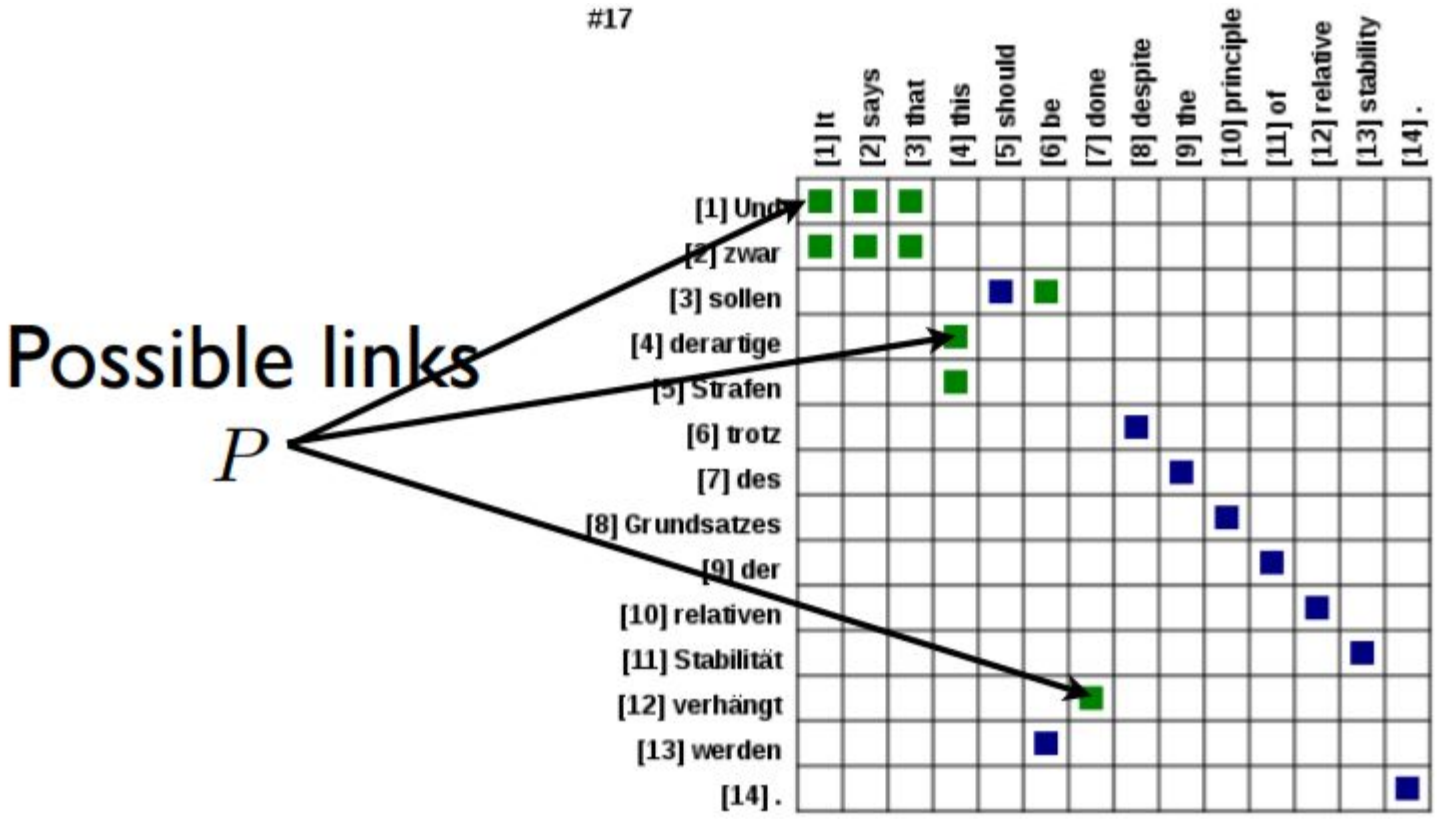
Evaluating Alignment Models

- How do we measure quality of a word-to-word model?
  - Method 1: use in an end-to-end translation system
    - Hard to measure translation quality
    - Option: human judges
    - Option: reference translations (NIST, BLEU)
    - Option: combinations (HTER)
    - Actually, no one uses word-to-word models alone as TMs
  - Method 2: measure quality of the alignments produced
    - Easy to measure
    - Hard to know what the gold alignments should be
    - Often does not correlate well with translation quality (like perplexity in LMs)

Alignment Error Rate



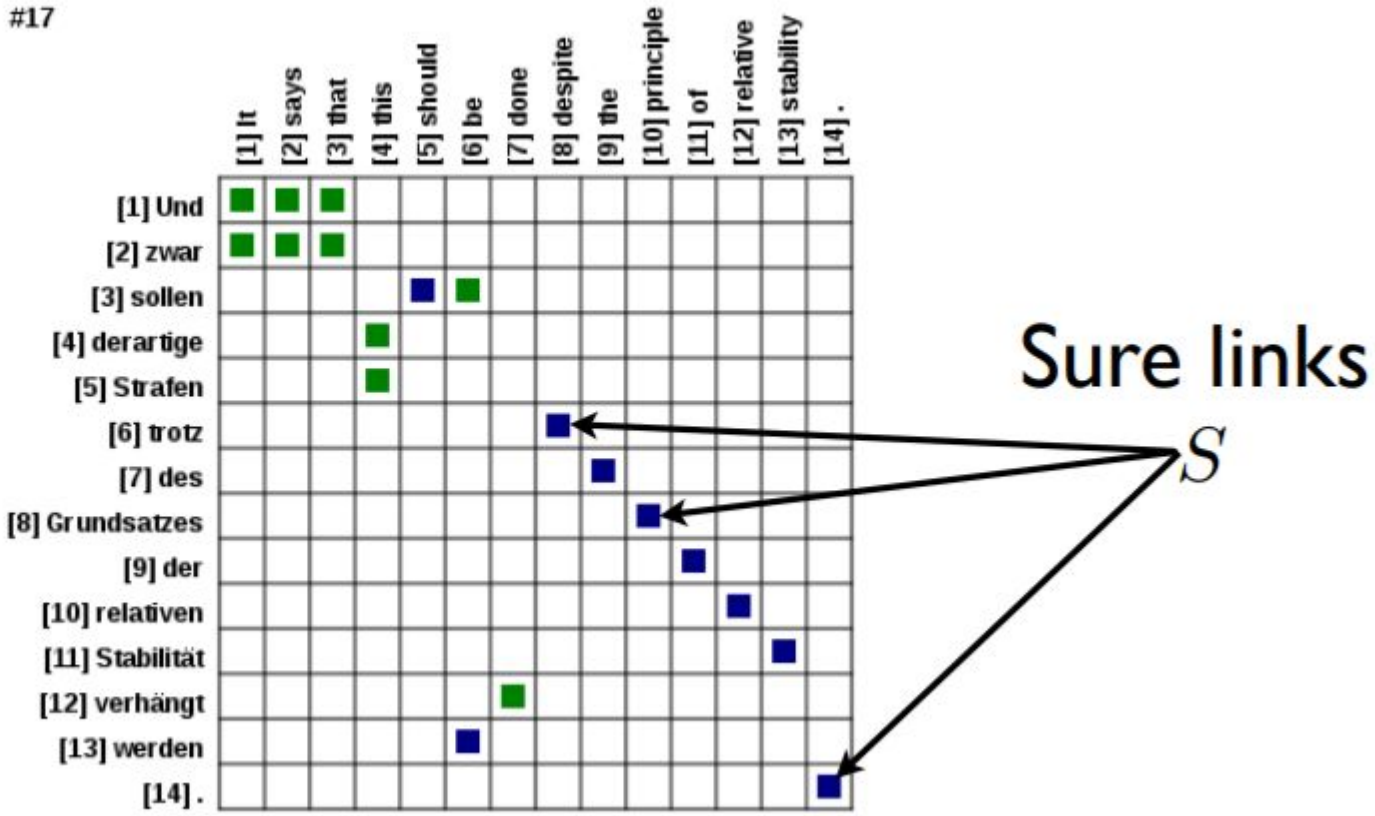
Alignment Error Rate



Alignment Error Rate

Possible links

P

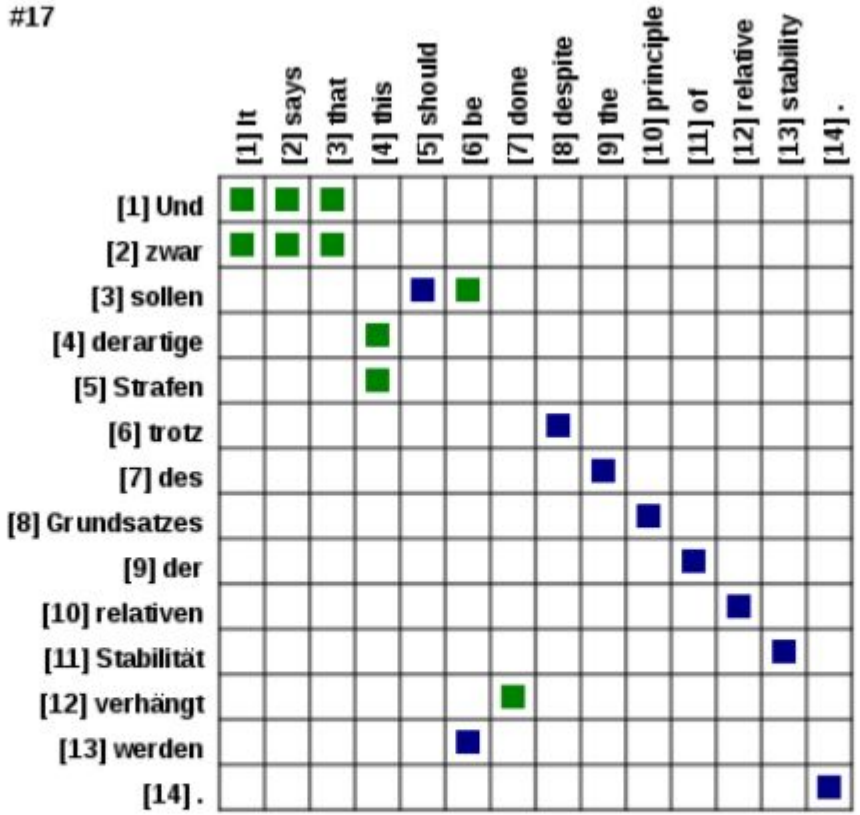


Sure links

S

Alignment Error Rate

Possible links
P

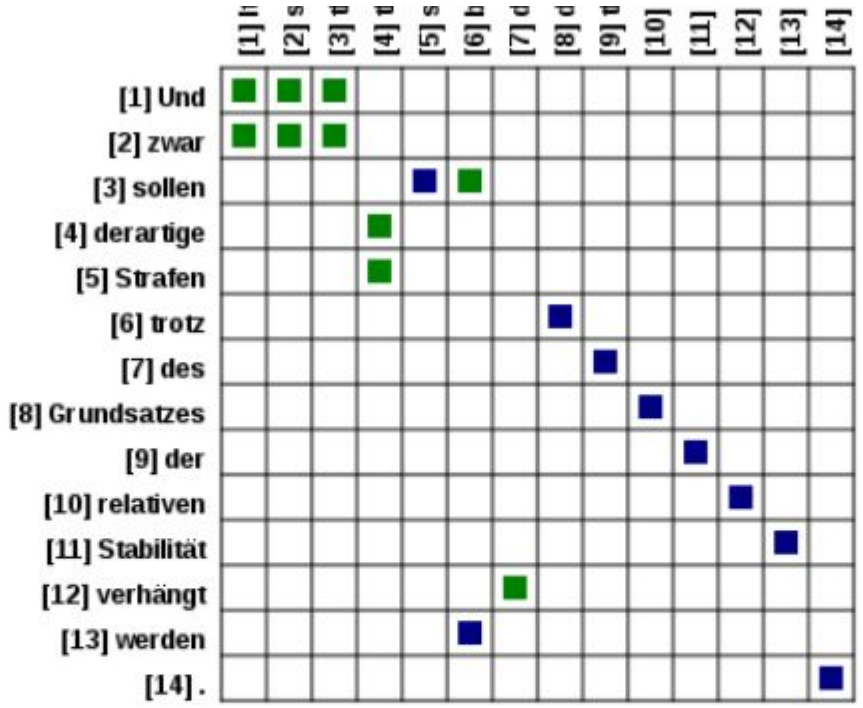


Sure links
S

$$\text{Precision}(A, P) = \frac{|P \cap A|}{|A|} \qquad \text{Recall}(A, S) = \frac{|S \cap A|}{|S|}$$

Alignment Error Rate

Possible links
P



Sure links
S

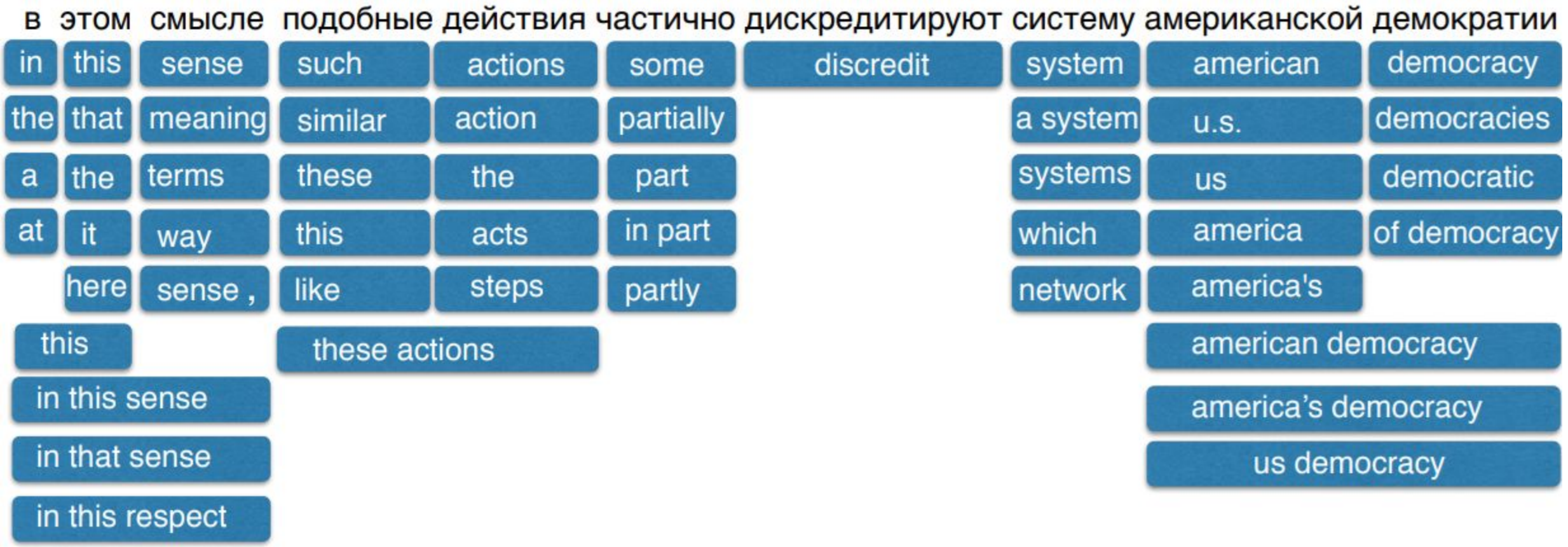
$$\text{Precision}(A, P) = \frac{|P \cap A|}{|A|} \qquad \text{Recall}(A, S) = \frac{|S \cap A|}{|S|}$$

$$\text{AER}(A, P, S) = 1 - \frac{|S \cap A| + |P \cap A|}{|S| + |A|}$$

Problems with Lexical Translation

- Complexity -- exponential in sentence length
- Weak reordering -- the output is not fluent
- Many local decisions -- error propagation

Phrase-Based Translation



$$P(e, \text{alignment} \mid f) = p_{segmentation} p_{\text{translation}} p_{reorderings}$$

Phrase-Based MT