Machine Translation

Chris Dyer
Language Technologies Institute
Machine Learning Department

MT Marathon 2013
September 9, 2013

Overview

- A Brief History of MT
- Modeling: Language & Translation
- Using MT
- Evaluating MT
- Coffee

MT Timeline

- 1940s - WW2 code breaking
- 1947 - Weaver letter outlining translation as a problem in cryptography
- 1954 - Georgetown Experiments showed "promise" of Russian-English MT
- 1966 - ALPAC report shifts funding to basic research in computational linguistics
- 1968 - MT company SYSTRAN founded (still in existence)
- 1970s - advances in formal languages and automata theory; development of *statistical speech recognition* techniques at IBM and Princeton
- 1993 - Weaver's model of translation prototyped by IBM; ***statistical revolution***
- 1999 - Open source reimplementation of IBM models
- 2000s - Major modeling advances, rediscovery of syntax, large scale funding
- 2006 - Open source Moses decoder development begins
- 2006 - Google Translate launches
- 2010 - SDL acquires Language Weaver
- 2013 - Prague MT Marathon!

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.***'

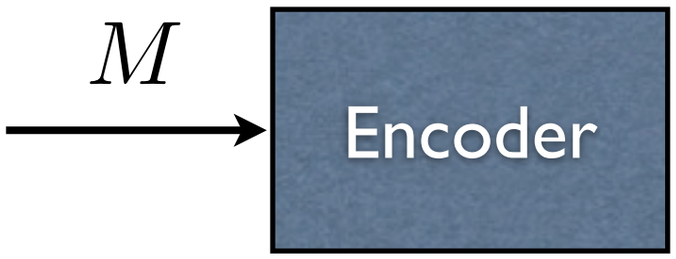
Warren Weaver to Norbert Wiener, March, 1947

Claude Shannon. "A Mathematical Theory of Communication" 1948.

M

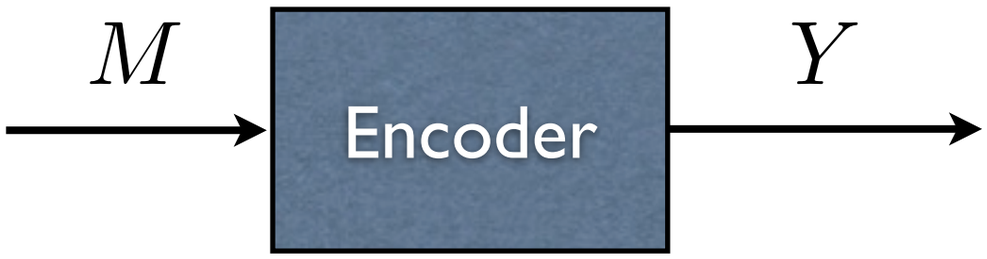
Message

Claude Shannon. "A Mathematical Theory of Communication" 1948.

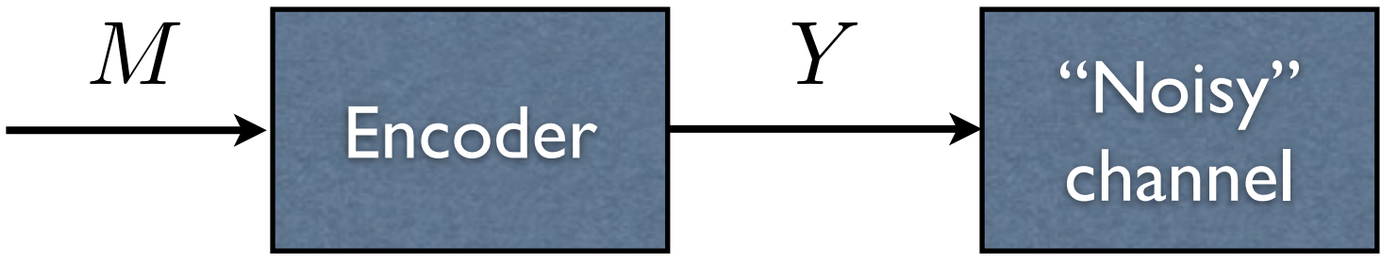


Message

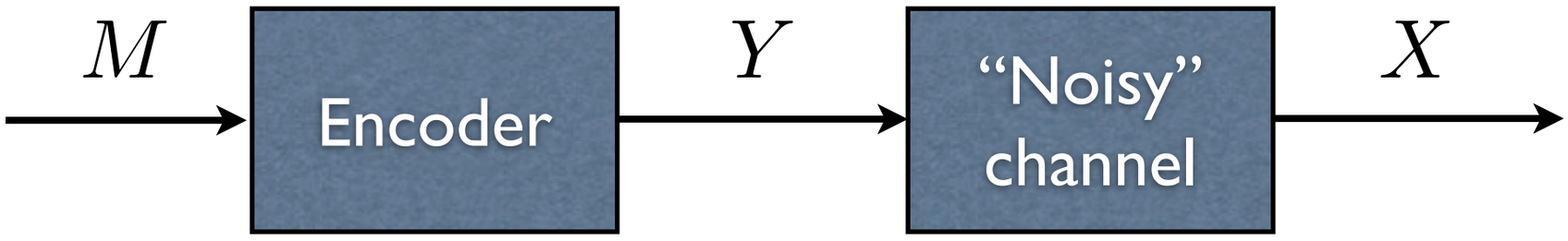
Claude Shannon. "A Mathematical Theory of Communication" 1948.



Claude Shannon. "A Mathematical Theory of Communication" 1948.



Claude Shannon. "A Mathematical Theory of Communication" 1948.



Claude Shannon. "A Mathematical Theory of Communication" 1948.

Claude Shannon. "A Mathematical Theory of Communication" 1948.

Claude Shannon. "A Mathematical Theory of Communication" 1948.

$$p(y) \quad p(x|y) \quad p(x)$$

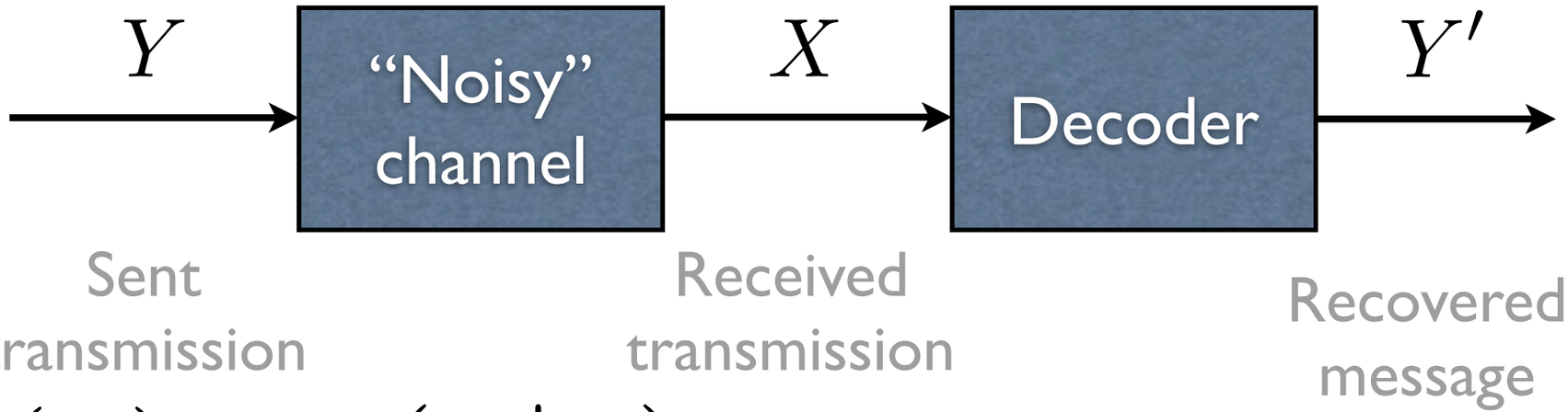
Claude Shannon. "A Mathematical Theory of Communication" 1948.

$$p(y) \quad p(x|y)$$

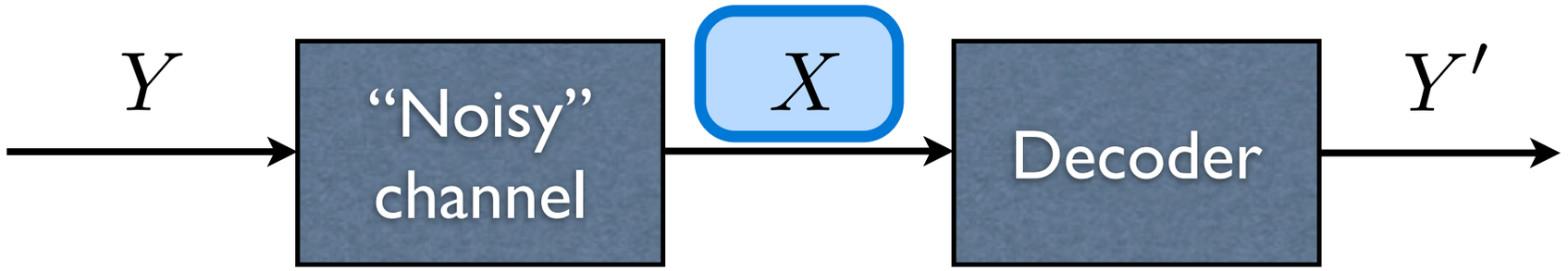
Shannon's theory tells us:

1) how much data you can send
2) the limits of compression
3) why your download is so slow
4) how to translate

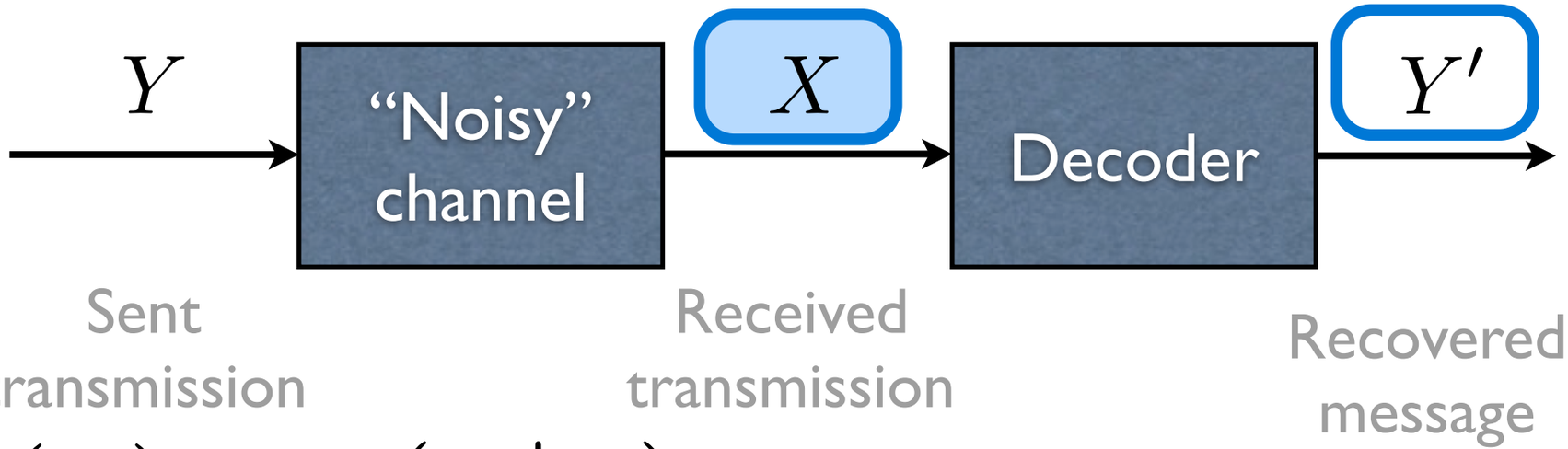
Claude Shannon. "A Mathematical Theory of Communication" 1948.


Y
"Noisy" channel
X
Decoder
Y'
Sent transmission
Received transmission
Recovered message

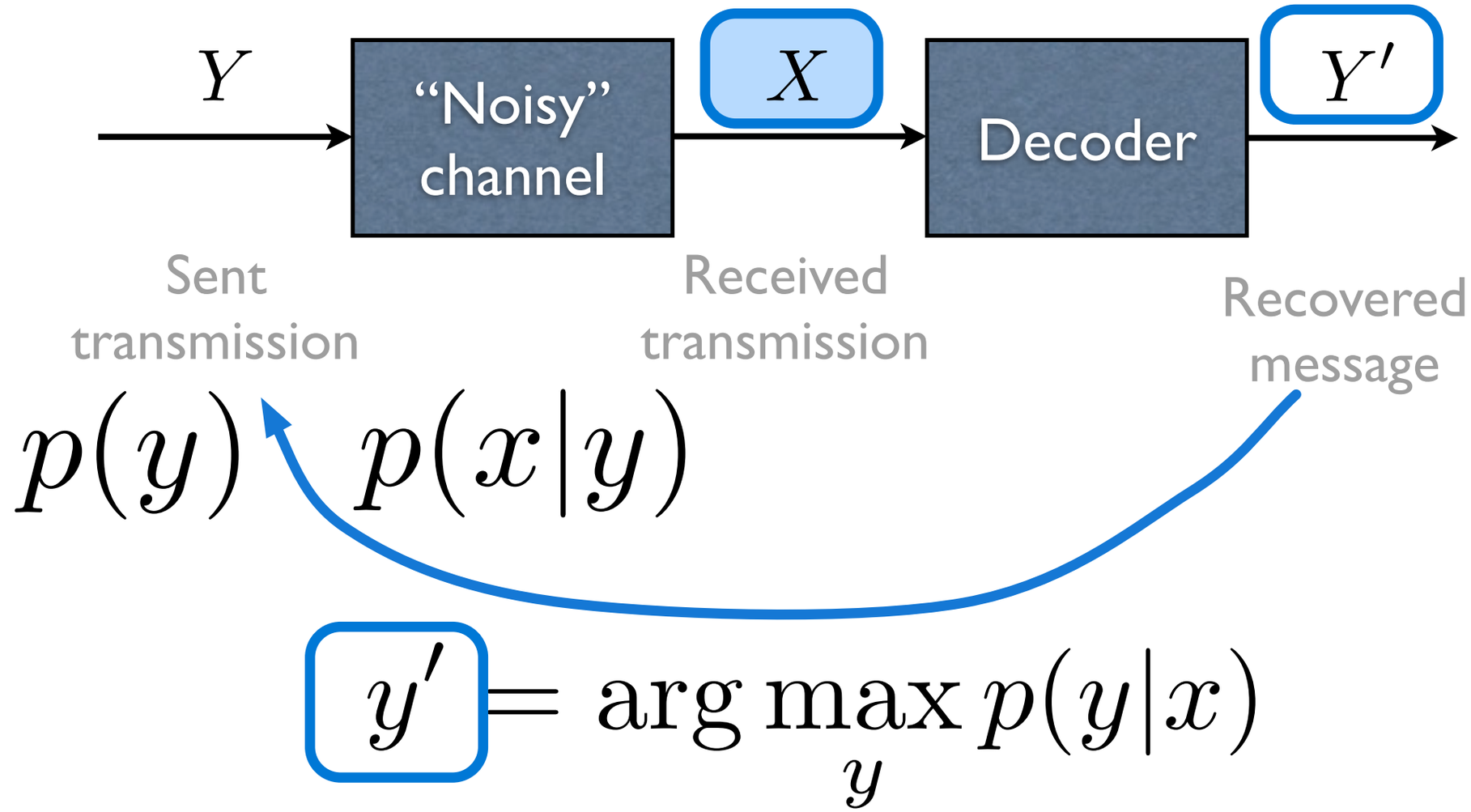
$$p(y) \quad p(x|y)$$

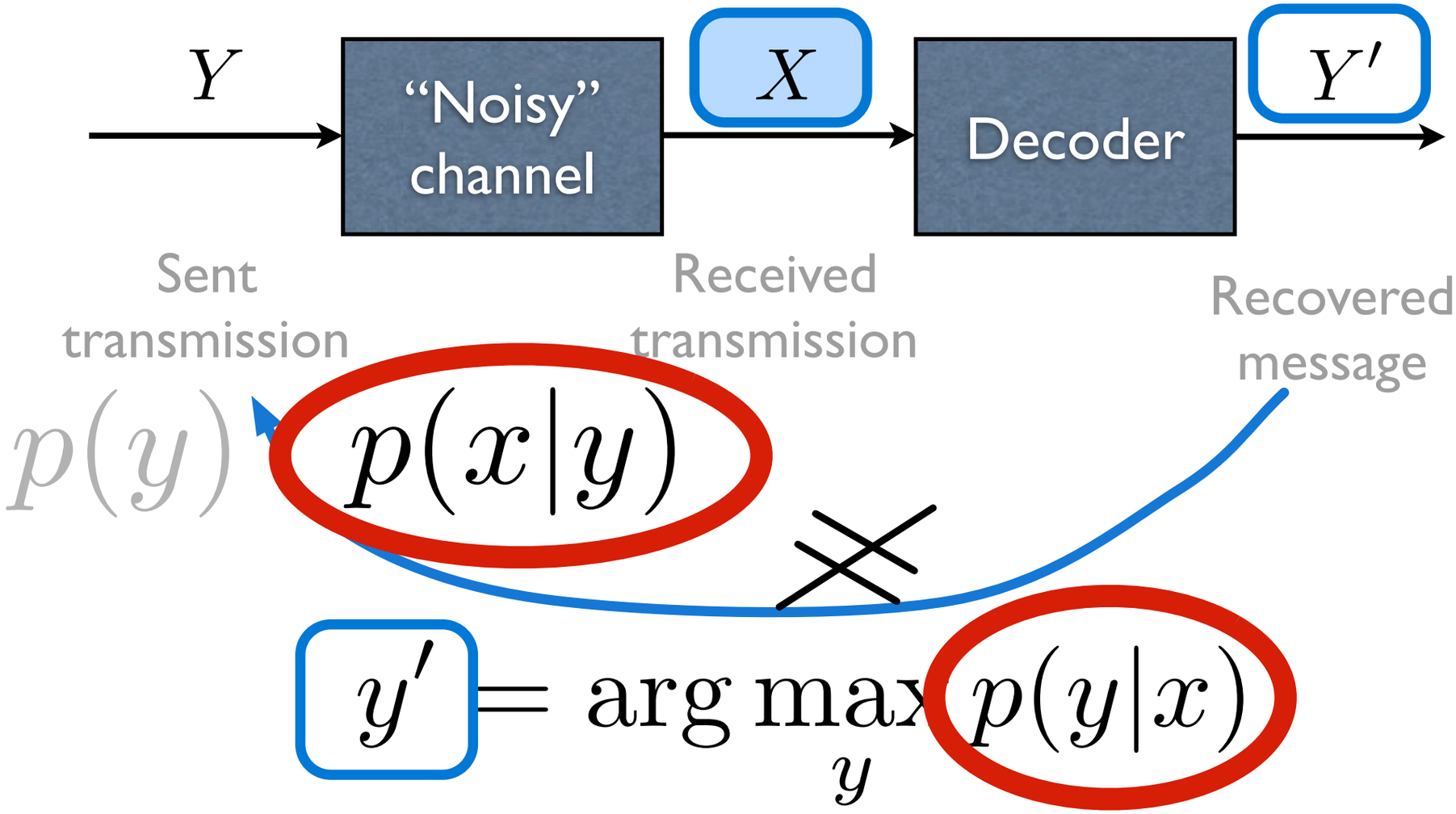

Y
"Noisy" channel
X
Decoder
Y'
Sent transmission
Received transmission
Recovered message

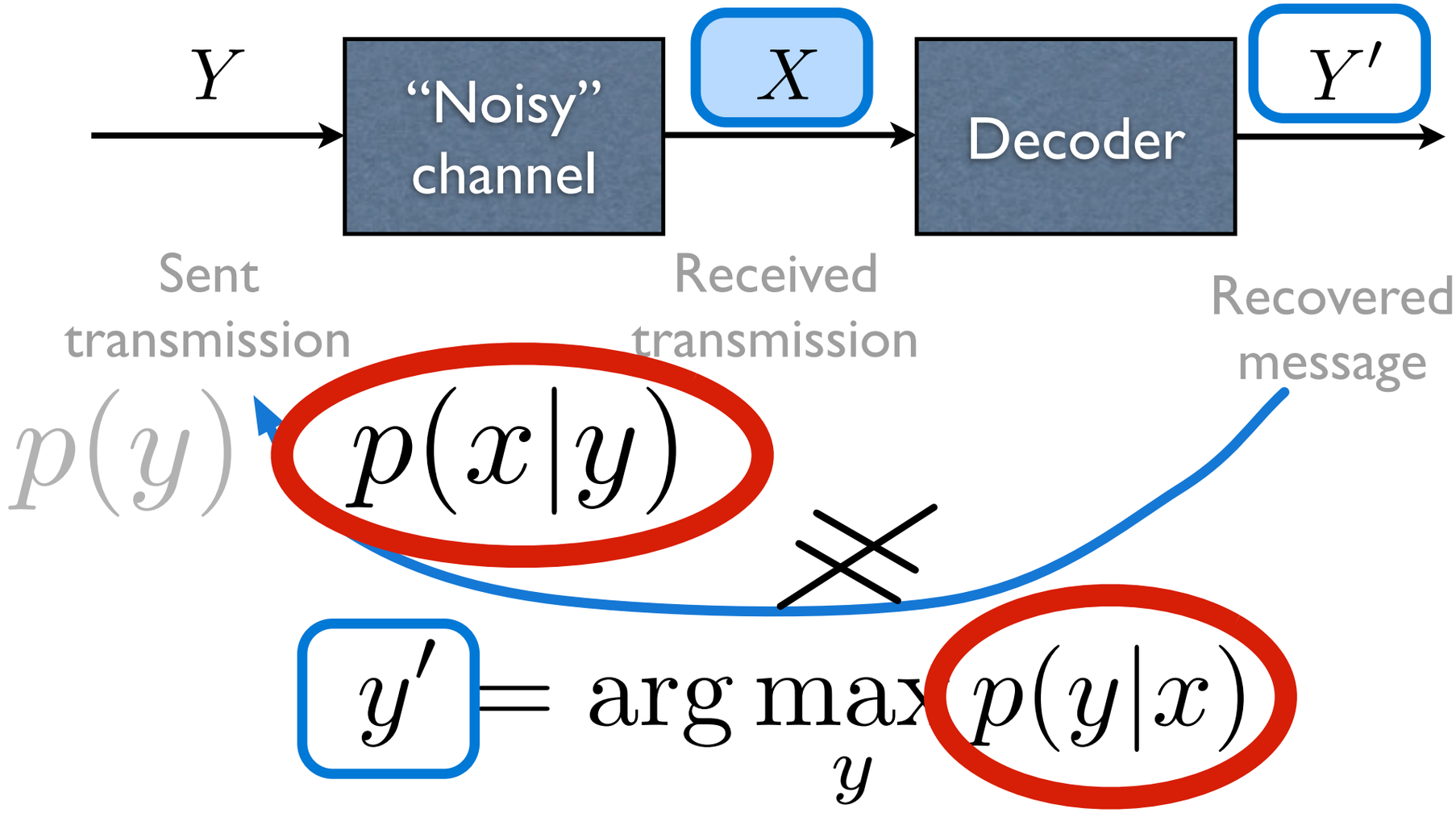
$$p(y) \quad p(x|y)$$


Y
"Noisy" channel
X
Decoder
Y'
Sent transmission
Received transmission
Recovered message

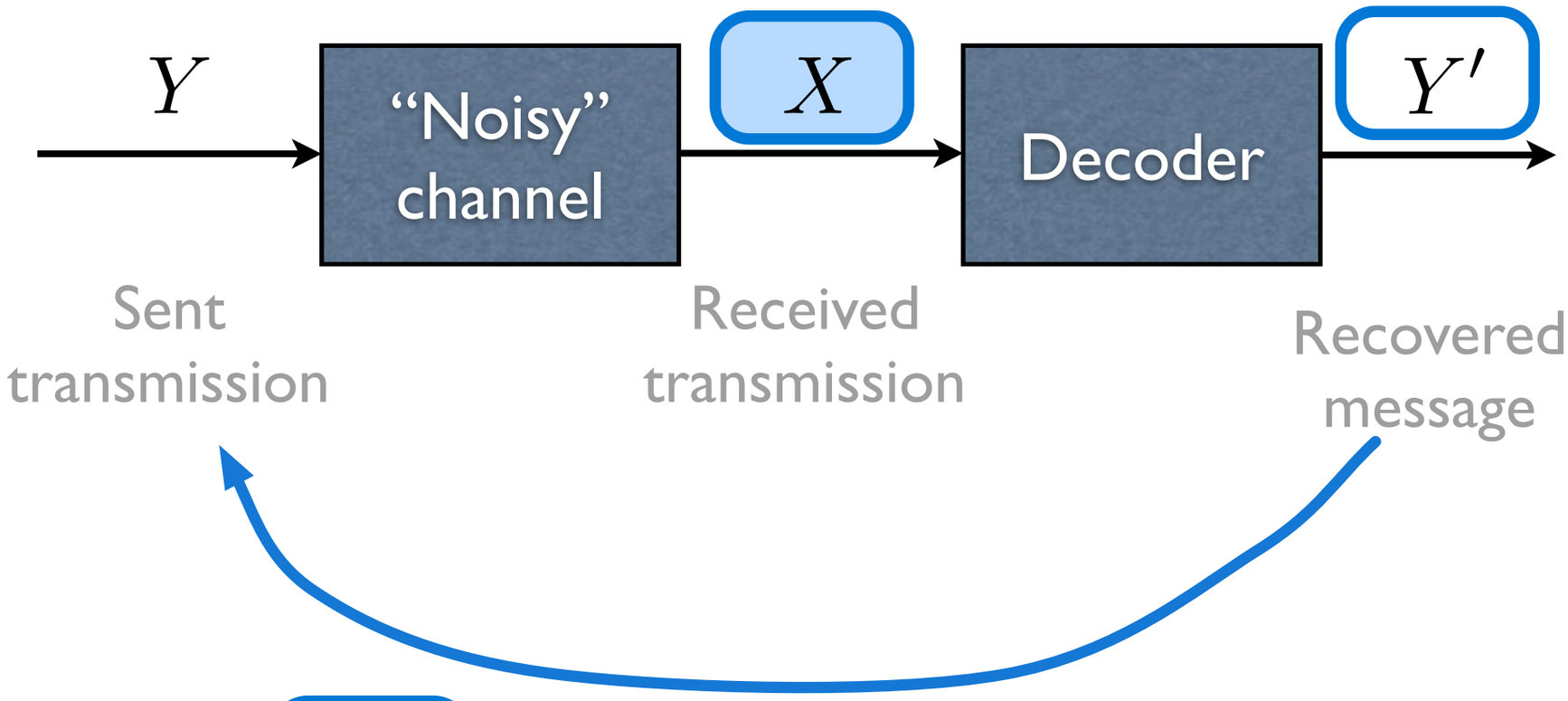
$$p(y) \quad p(x|y)$$


Y
"Noisy" channel
X
Decoder
Y'
Sent transmission
Received transmission
Recovered message
$p(y)$
$p(x|y)$
$y' = \arg\max_y p(y|x)$


Y
"Noisy" channel
X
Decoder
Y'
Sent transmission
Received transmission
Recovered message
$p(y)$
$p(x|y)$
$y' = \arg\max_y p(y|x)$

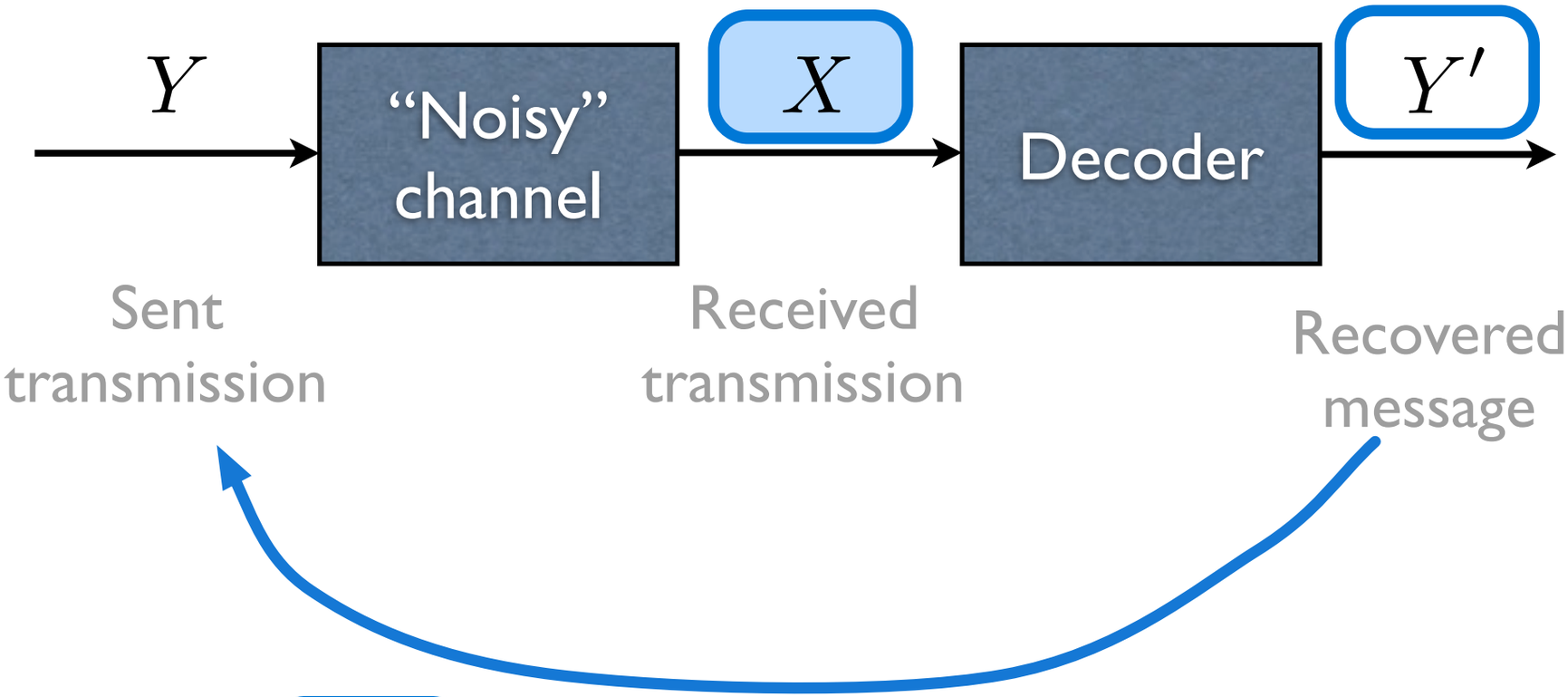

Y
"Noisy" channel
X
Decoder
Y'
Sent transmission
Received transmission
Recovered message
$p(y)$
$p(x|y)$
$y' = \arg\max_y p(y|x)$

I can help.



$$y' = \arg\max_y p(y|x)$$

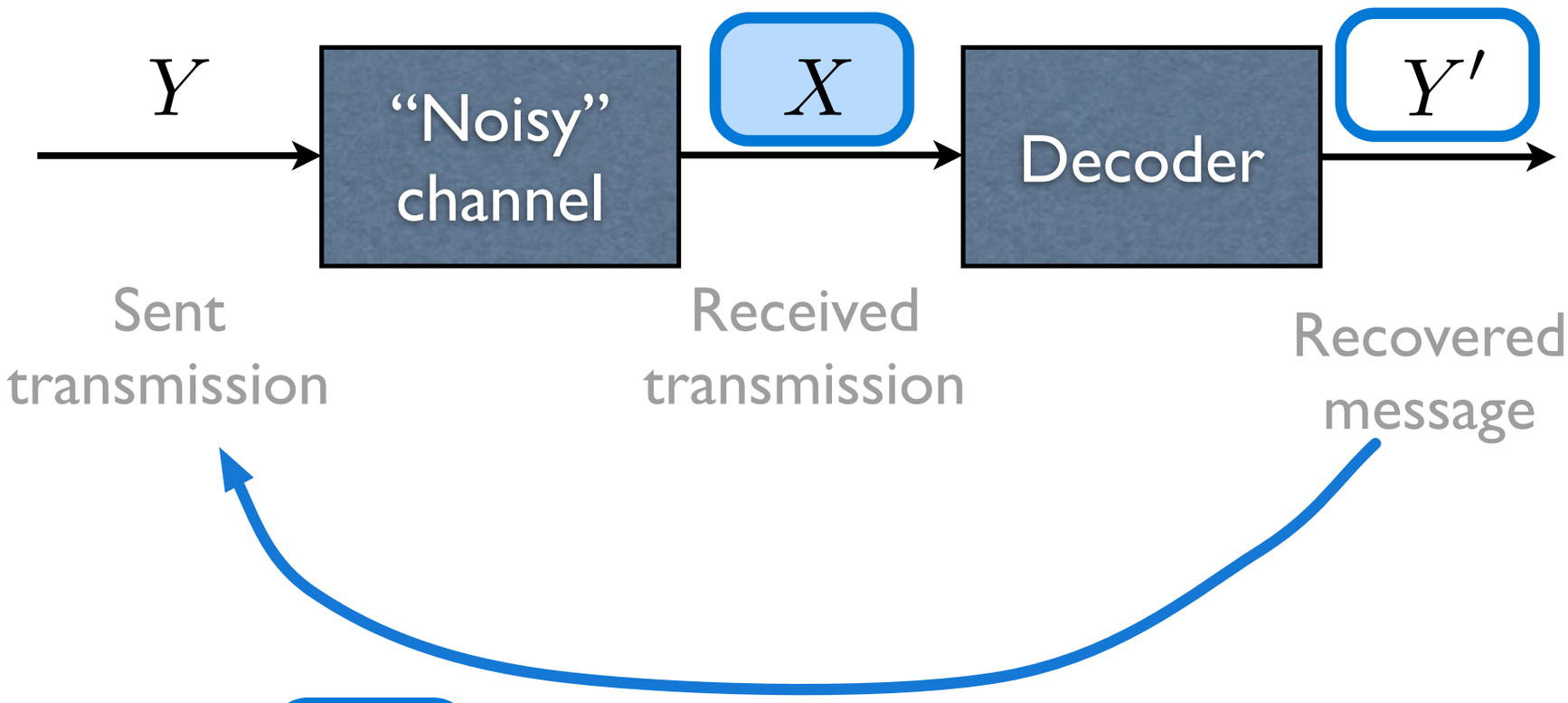
$$= \arg\max_y \frac{p(x|y)p(y)}{p(x)}$$



$$y' = \arg\max_y p(y|x)$$

$$= \arg\max_y \frac{p(x|y)p(y)}{p(x)}$$

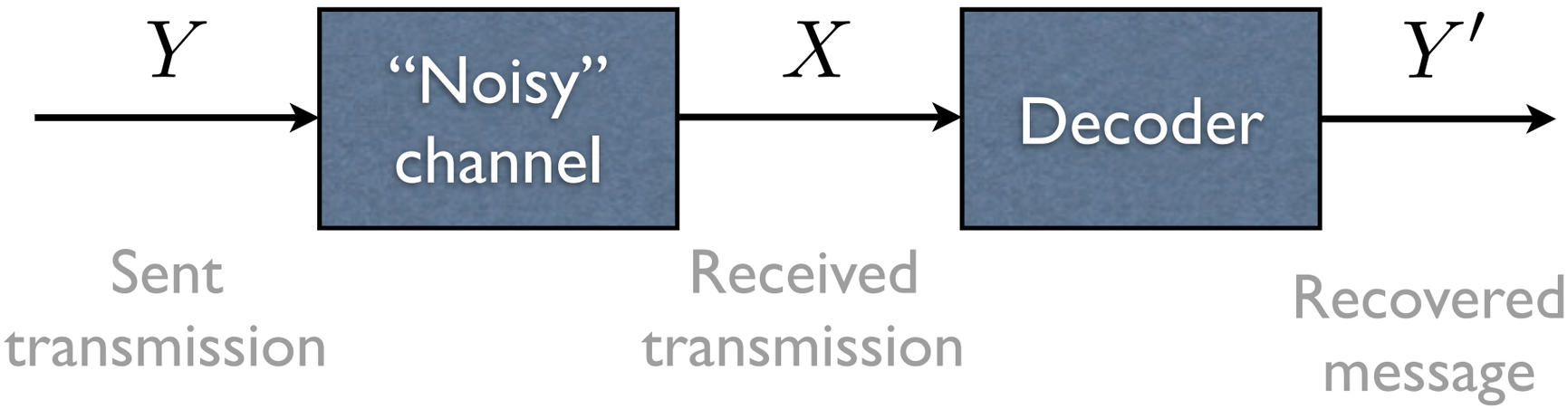
Denominator doesn't depend on y.



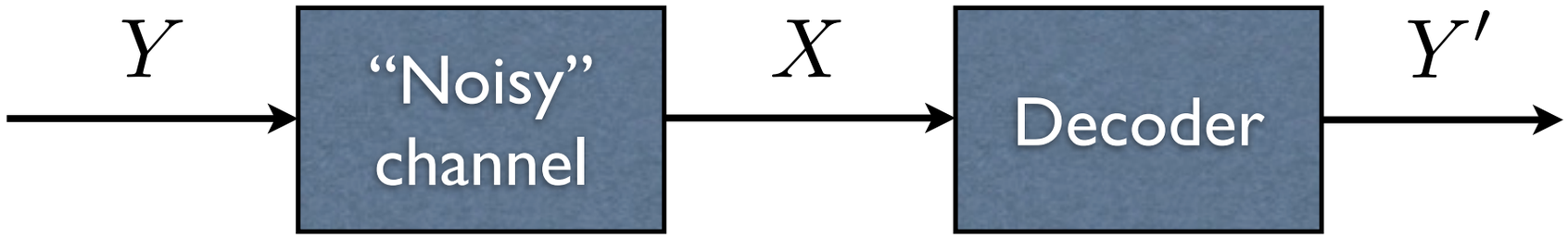
$$y' = \arg\max_y p(y|x)$$

$$= \arg\max_y \frac{p(x|y)p(y)}{p(x)}$$

$$= \arg\max_y p(x|y)p(y)$$

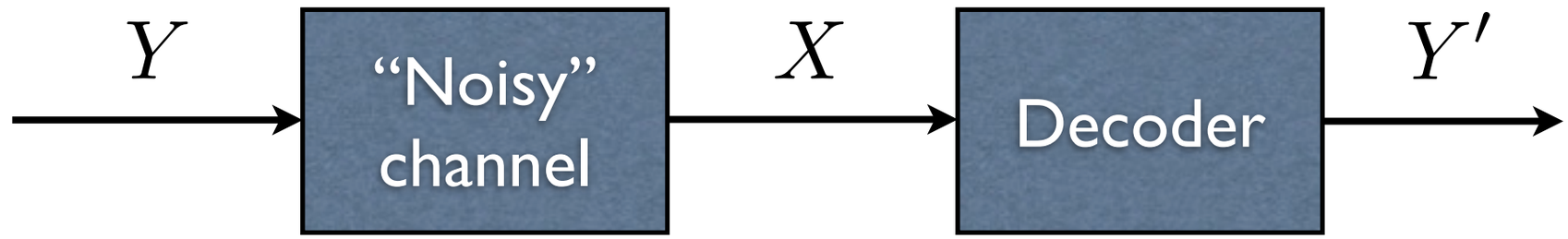


$$y' = \arg\max_y p(x|y)p(y)$$



$$y' = \arg\max_{y} p(x|y)p(y)$$

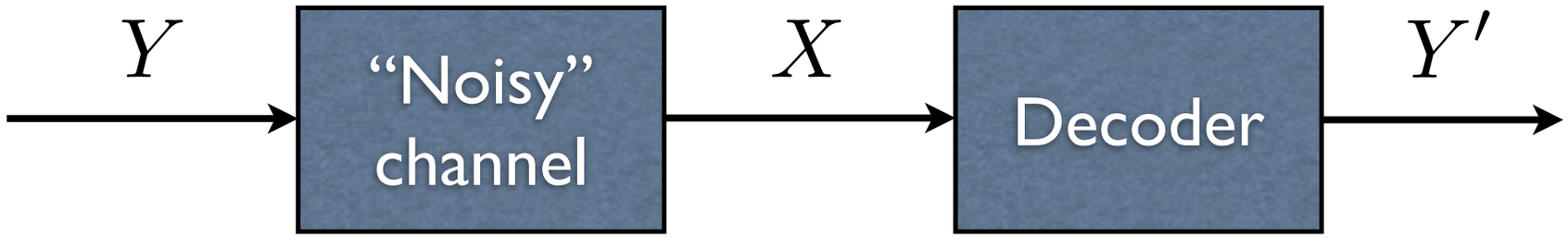
$$\mathbf{e}' = \arg\max_{\mathbf{e}} p(\mathbf{f}|\mathbf{e})p(\mathbf{e})$$



$$y' = \arg\max_y p(x|y)p(y)$$

$$\mathbf{e}' = \arg\max_{\mathbf{e}} p(\mathbf{f}|\mathbf{e})p(\mathbf{e})$$

translation model



Sent transmission

Received transmission

Recovered message

English **Csesky** **English'**

$$y' = \arg\max_y p(x|y)p(y)$$

$$\mathbf{e}' = \arg\max_{\mathbf{e}} p(\mathbf{f}|\mathbf{e})p(\mathbf{e})$$

translation model

language model

Division of labor

- **Translation model**
  - translation *back* into the source
  - learned from (source, target) translations
    - **adequacy** of translation
- **Language model**
  - probability of the output sentence
  - learned from *any* target language corpus
  - **fluency** of translation

Intuition for Division of Labor

- Better use of data
  - We have parallel data
  - We also have (lots of) parallel data
    - ... use both
- Use weaker translation models
  - Language modeling is hard ...
  - Translation modeling is language modeling ++
- Language models are used in many applications

The Big Question in MT

The Big Question in MT

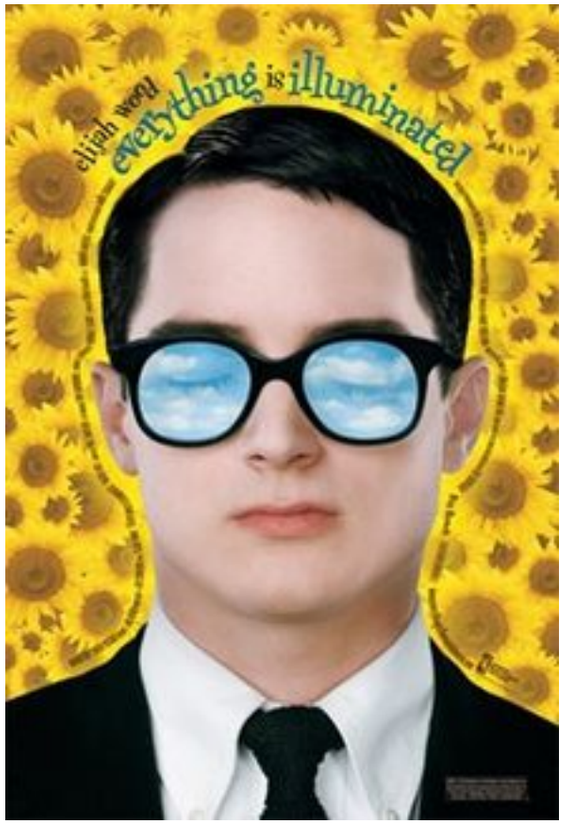
- **How do we *design* language and translation models?**

The Big Question in MT

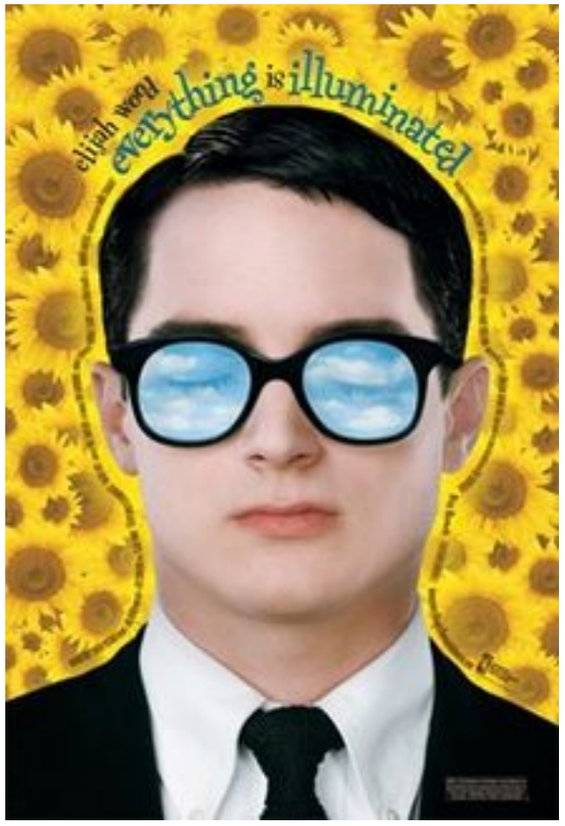
- **How do we *design* language and translation models?**
- Considerations
  - Is the model ***correct***?
  - Is prediction (*inference*) ***tractable***?
  - Is there data to learn the parameters?

Language Models

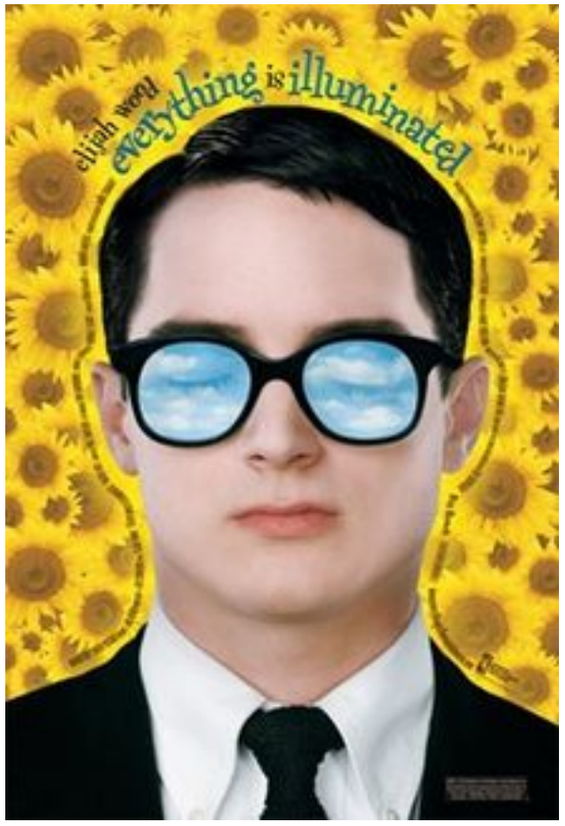
- What is the **probability** of a sentence (in a particular language)?
  - There are an *infinite number* of grammatical sentences in any language
  - This is the era of Big Data...but infinity is bigger
- For machine translation
  - Naive model (model language)
  - Naive parameterization (no relationship between words like **walk walked walking walks**)
  - Clever estimation of parameters (Lecture later this week)



My legal name is Alexander Perchov. But all of my many friends dub me Alex, because that is a more flaccid-to-utter version of my legal name.



My legal name is Alexander Perchov. But all of my many friends dub me Alex, because that is a more flaccid-to-utter version of my legal name. Mother dubs me Alexi-stop-spleening-me!, because I am always spleening her.



My legal name is Alexander Perchov. But all of my many friends dub me Alex, because that is a more flaccid-to-utter version of my legal name. Mother dubs me Alexi-stop-spleening-me!, because I am always spleening her. If you want to know why I am always spleening her, it is because I am always elsewhere with friends, and disseminating so much currency, and performing so many things that can spleen a mother.

Probability Models of Text

- Sequence of symbols (bytes, letters, characters, morphemes, words, ...)
  - Let Σ denote the set of symbols
- Lots of possible sequences (Σ^* is infinitely large!)
- Probability distributions over Σ^*?

Probability Models of Text

- How are probability models built?
  - Make some independence assumptions

$$P(x, y) = P(x) \times P(y) \iff x, y \text{ are independent}$$

  - Make assumptions about the (conditional) distributions of the events
  - Estimate parameters from a sample of data

I ^ want

I want a

I want a flight

I want a flight to

I want a flight to Prague

I want a flight to Prague STOP

I want a flight to Prague STOP

English:

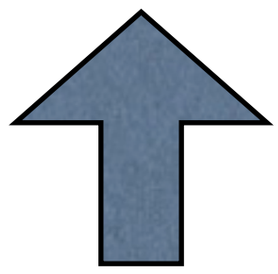
I have a bad headache

Chukchi (Siberian language):

English:

I have a bad headache

Chukchi (Siberian language):



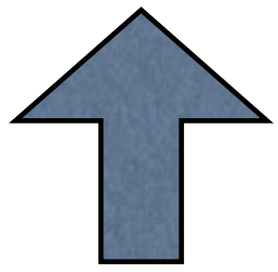
Aggultinative and polysynthetic languages have rich word-formation processes.

English:

I have a bad headache

Chukchi (Siberian language):

Təmeyŋəlevtpəγtərkən



Aggultinative and polysynthetic languages have rich word-formation processes.

Translation Models

- Is a string of words **e** a meaning-preserving translation of a string of words **f**?
- Conditional model p(**f** | **e**)
- Challenges
  - There are an infinite number of sentences
  - How do we learn parameters of the models?

In der Innenstadt explodierte eine Autobombe

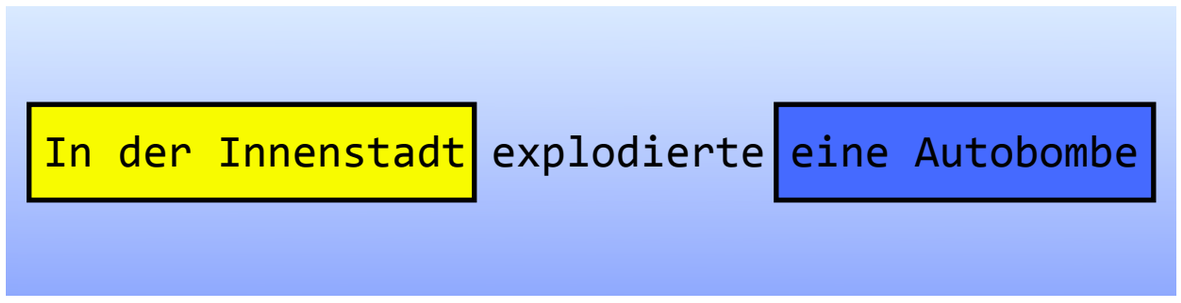
- **Make independence assumptions**

In der Innenstadt explodierte eine Autobombe

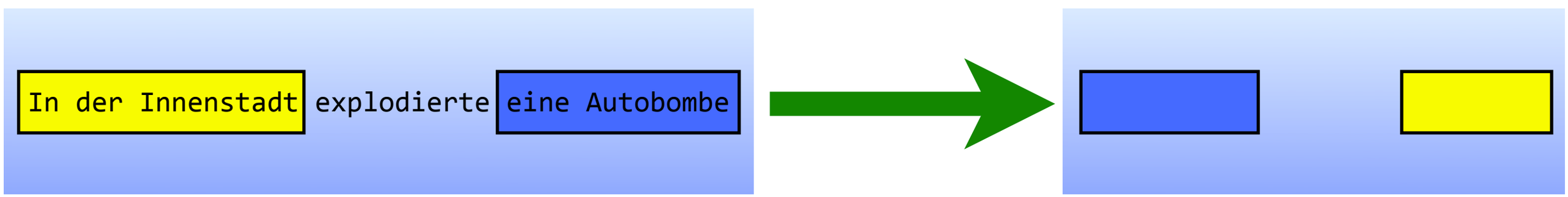
- **Make independence assumptions**
  - Permute the source words into the target language order

In der Innenstadt explodierte eine Autobombe

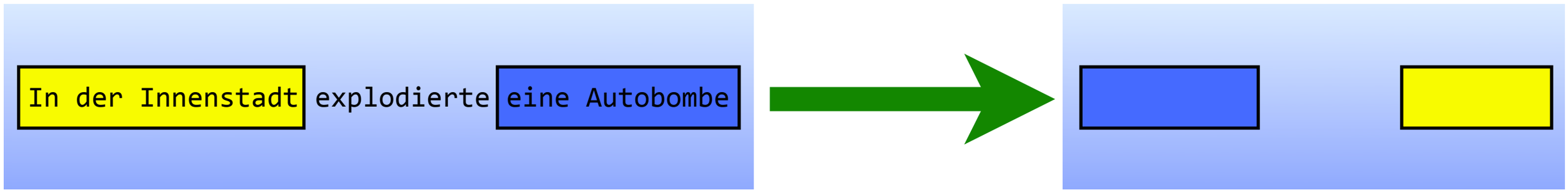
- **Make independence assumptions**
  - Permute the source words into the target language order



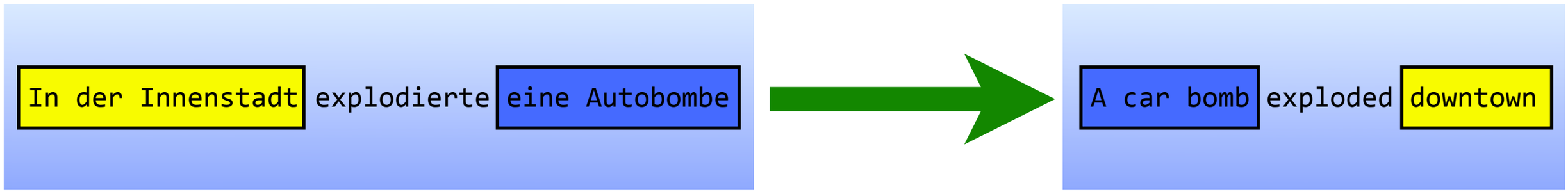
- **Make independence assumptions**
  - Permute the source words into the target language order



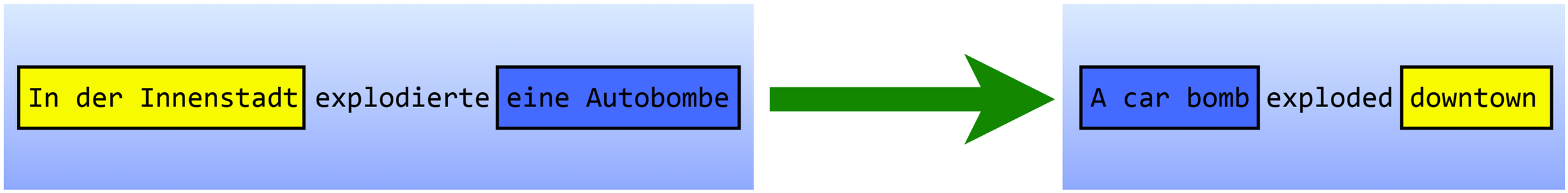
- **Make independence assumptions**
  - Permute the source words into the target language order
  - Pick translations for individual words / phrases
    - Probability of particular translations
    - Look at source context
    - Ensure that the output is fluent & idiomatic



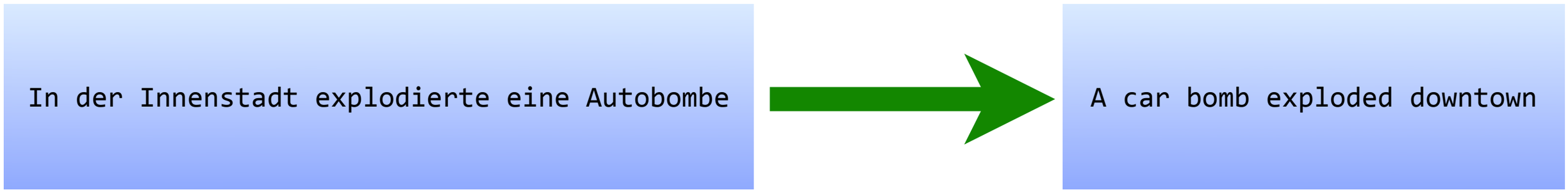
- **Make independence assumptions**
  - Permute the source words into the target language order
  - Pick translations for individual words / phrases
    - Probability of particular translations
    - Look at source context
    - Ensure that the output is fluent & idiomatic

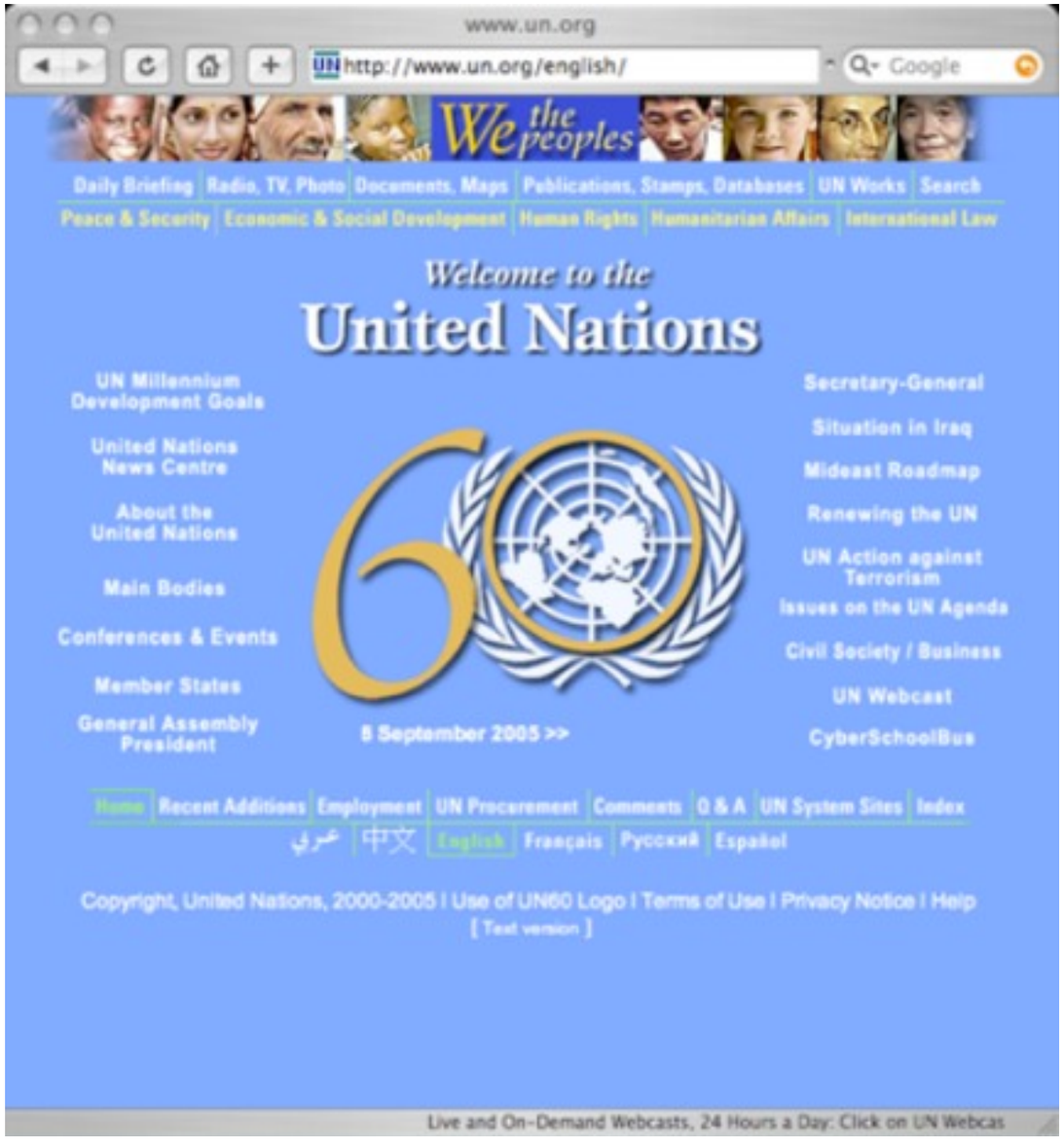


- **Make independence assumptions**
  - Permute the source words into the target language order
  - Pick translations for individual words / phrases
    - Probability of particular translations
    - Look at source context
    - Ensure that the output is fluent & idiomatic
- **Computational challenges**
  - Searching all word permutations is NP-hard
  - Massive numbers of translation alternatives

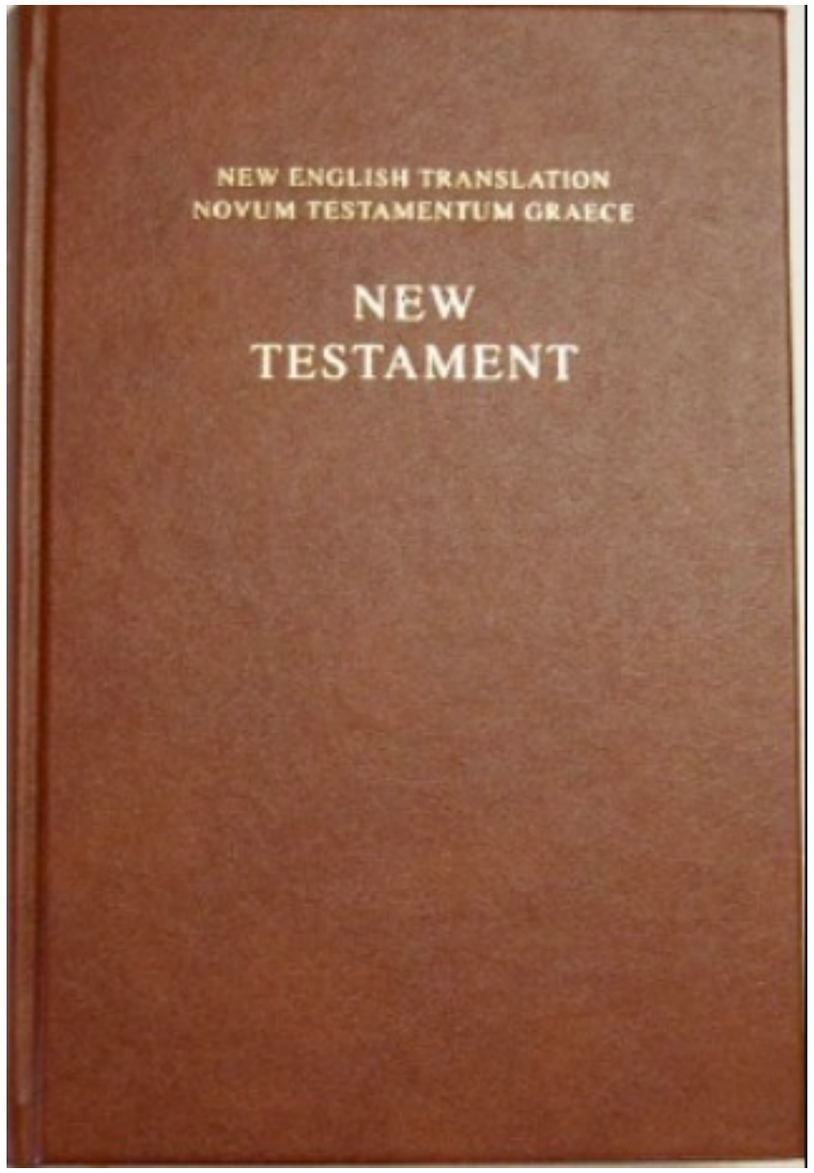


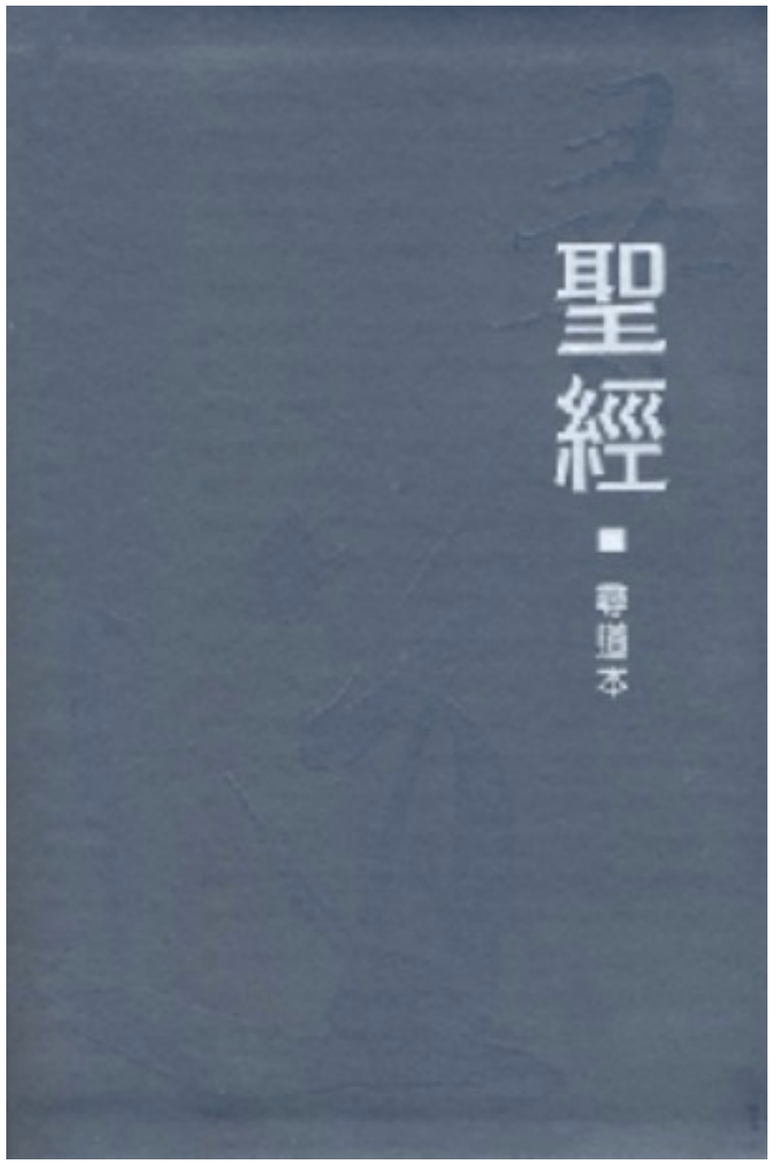
- **String-to-string translation**
  - Became popular in the 1990's with *statistical MT*
  - State-of-the-art for many (most?) language pairs
    - Especially: Closely related language pairs
    - Especially: Typologically similar language pairs
    - ~ Google Translate / Bing Translator
  - Limitations
    - Independence assumptions are wrong (too strong *and* too weak)
    - No structural information available to improve modeling
  - **Models can be learned directly from *parallel data***

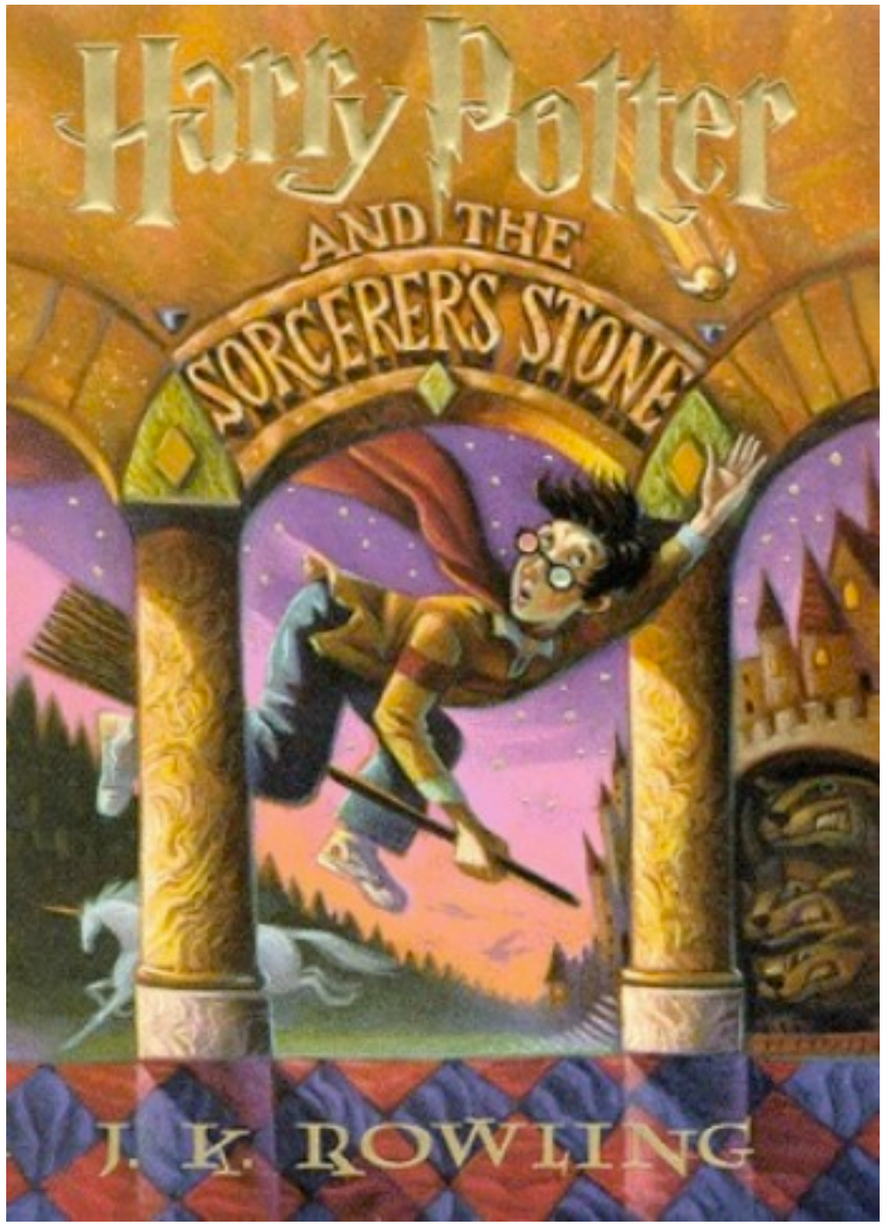


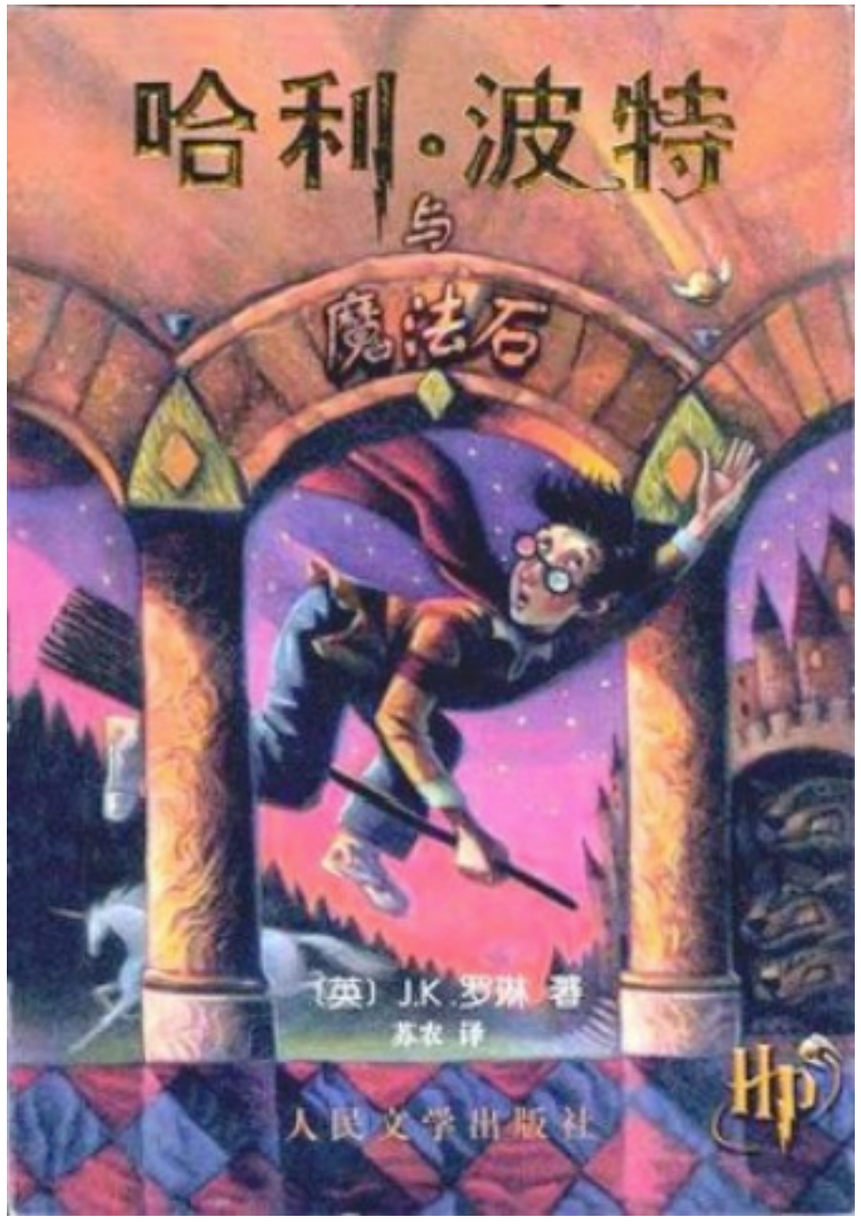

www.un.org
http://www.un.org/english/
Google
We the peoples
Daily Briefing | Radio, TV, Photo | Documents, Maps | Publications, Stamps, Databases | UN Works | Search
Peace & Security | Economic & Social Development | Human Rights | Humanitarian Affairs | International Law
Welcome to the
United Nations
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 UN60 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]


NEW ENGLISH TRANSLATION
NOVUM TESTAMENTUM GRAECE
NEW
TESTAMENT

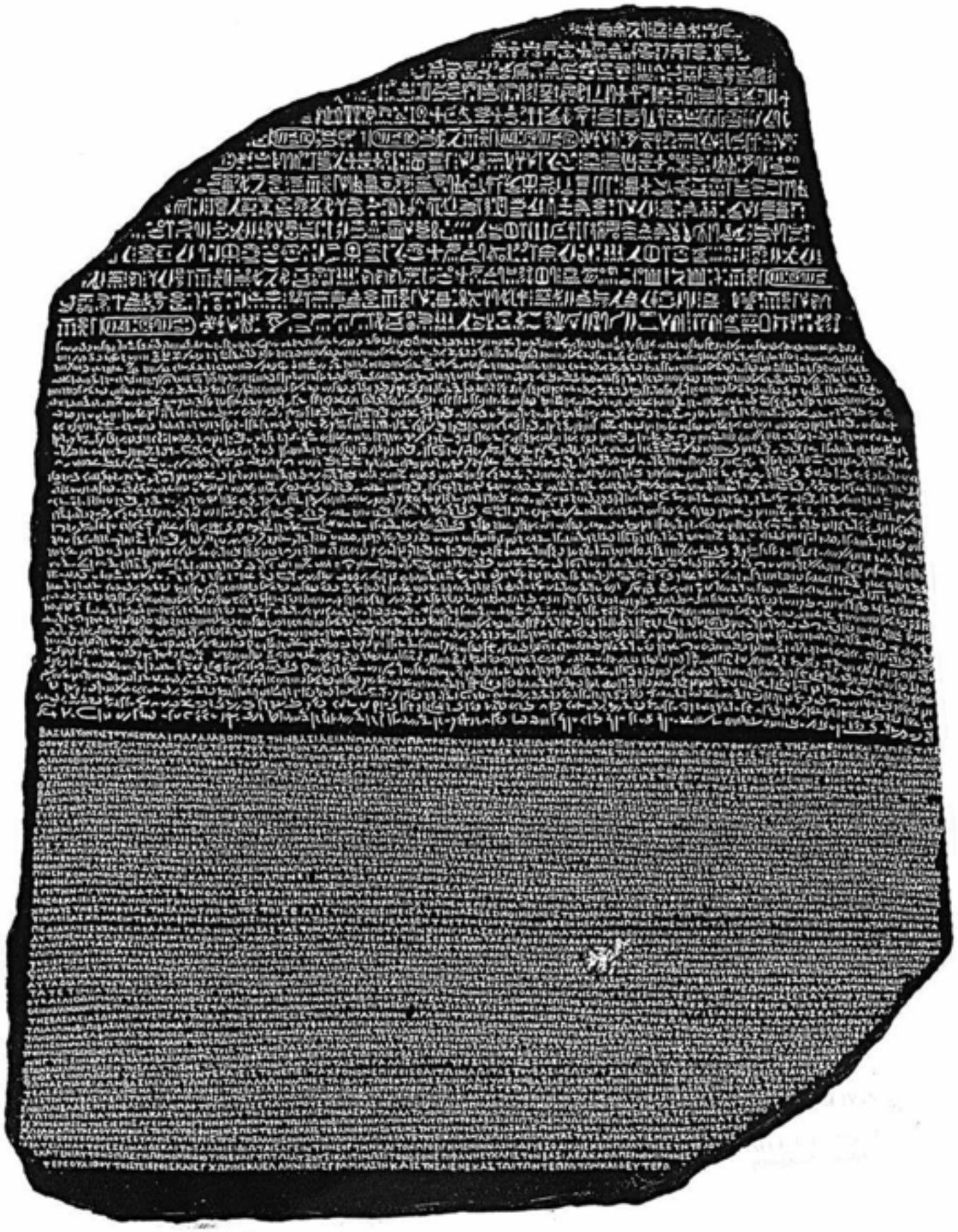

聖經


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
Hp

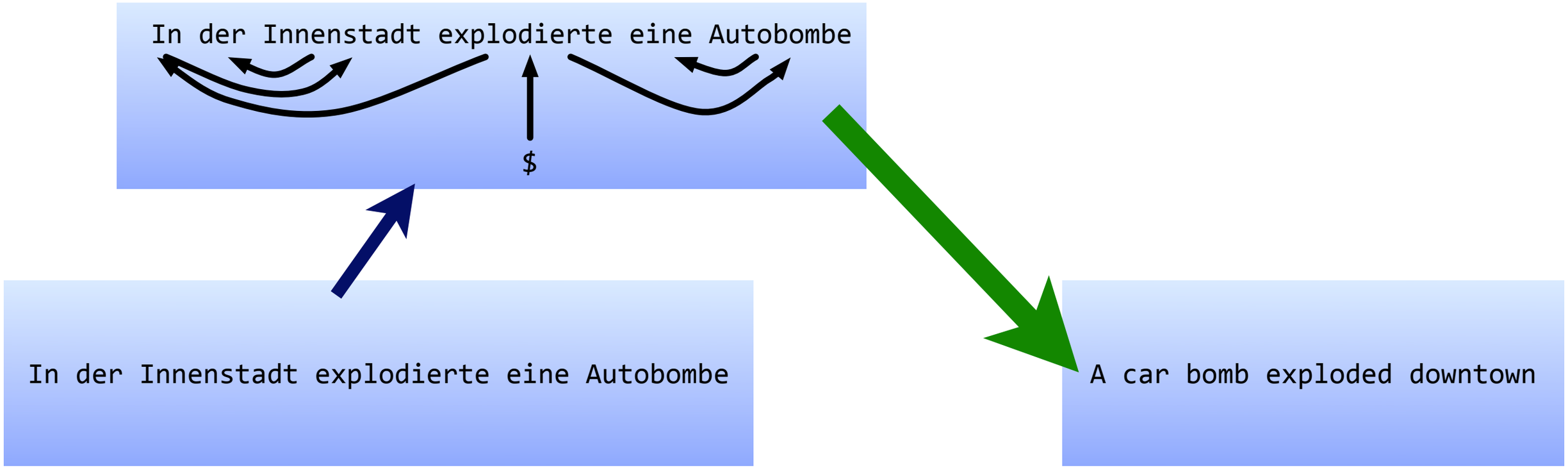
Egyptian

Greek

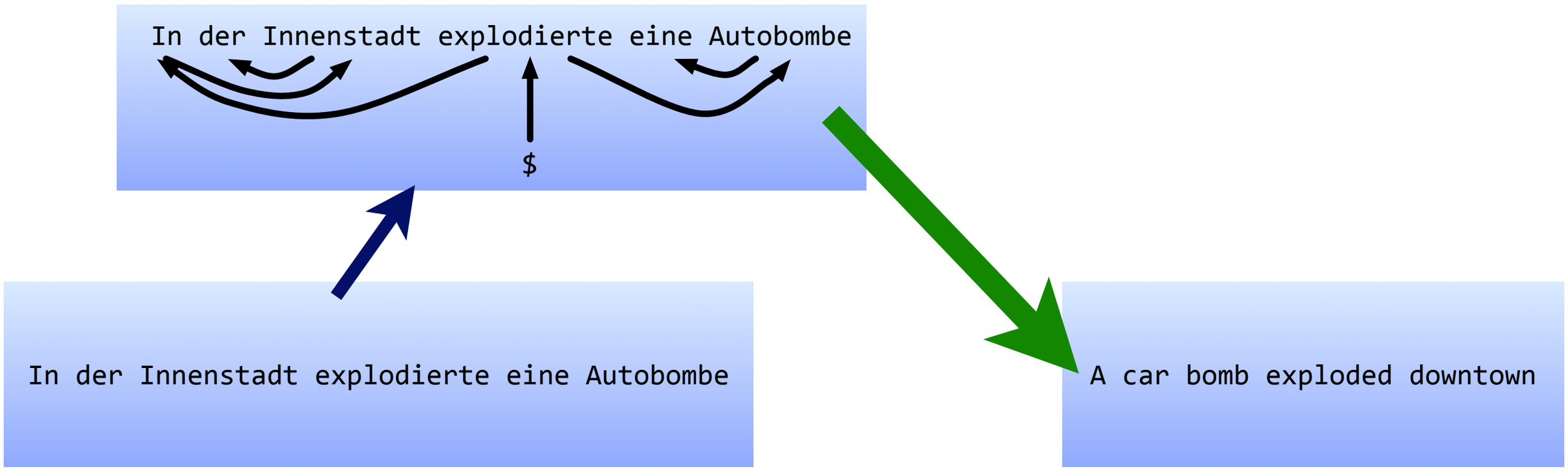


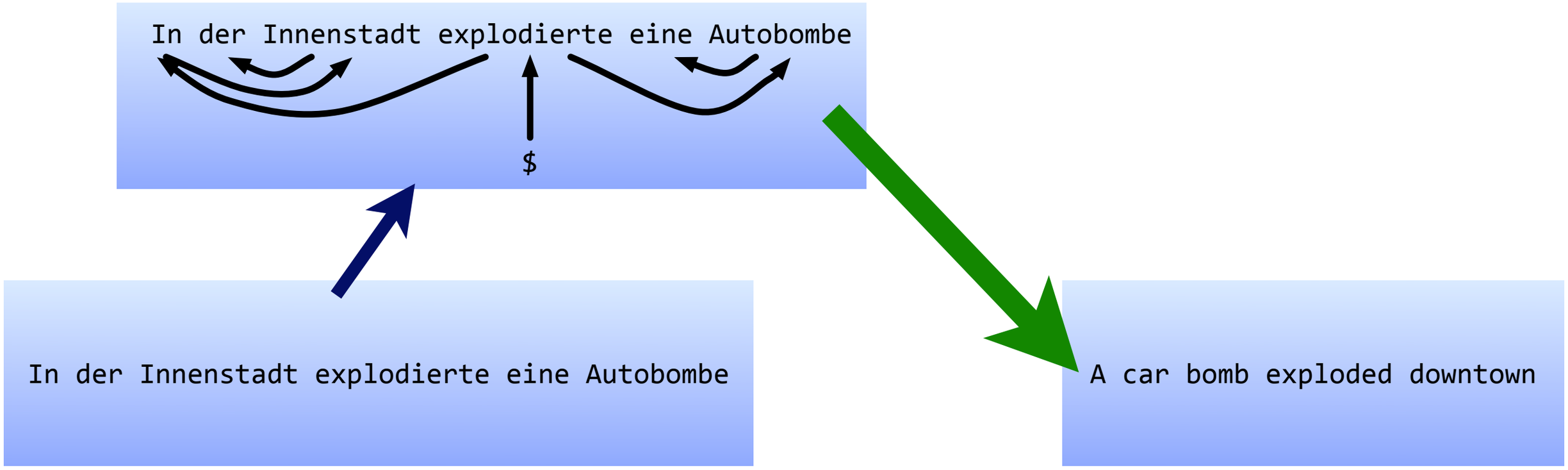
In der Innenstadt explodierte eine Autobombe

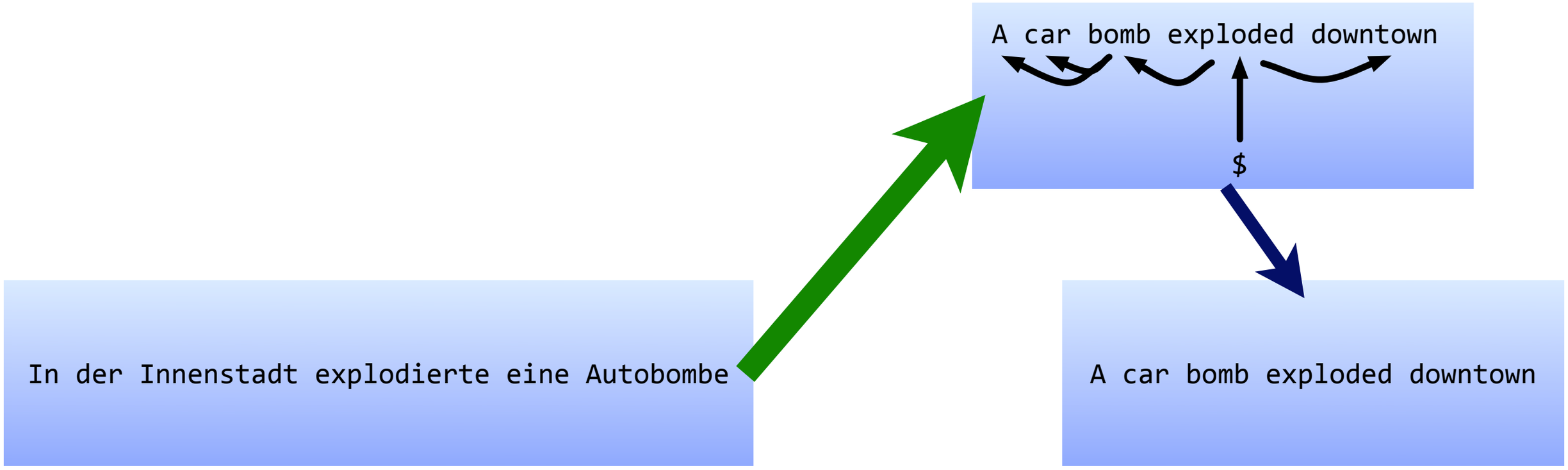
A car bomb exploded downtown


In der Innenstadt explodierte eine Autobombe
$
In der Innenstadt explodierte eine Autobombe
A car bomb exploded downtown

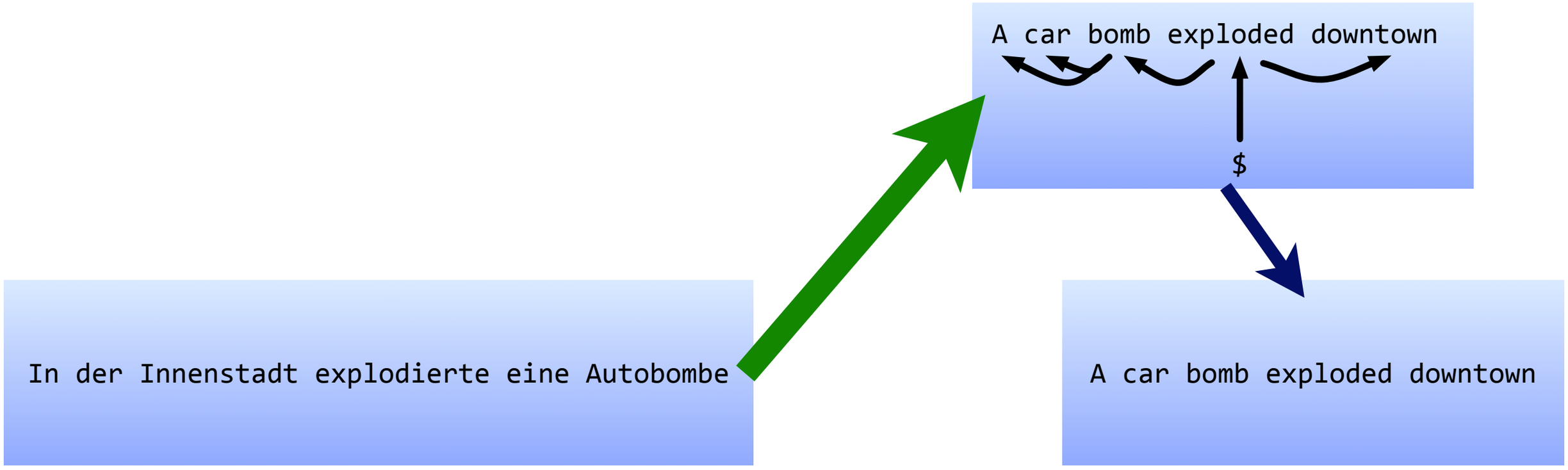
- **Tree-to-string translation**
  - Syntactic *analysis* of source (parse)
  - *Transfer* from tree to string
  - Source trees have some benefits
    - Proxy for semantic relationships
    - Syntax is a natural source of reordering constraints




In der Innenstadt explodierte eine Autobombe
$
In der Innenstadt explodierte eine Autobombe
A car bomb exploded downtown

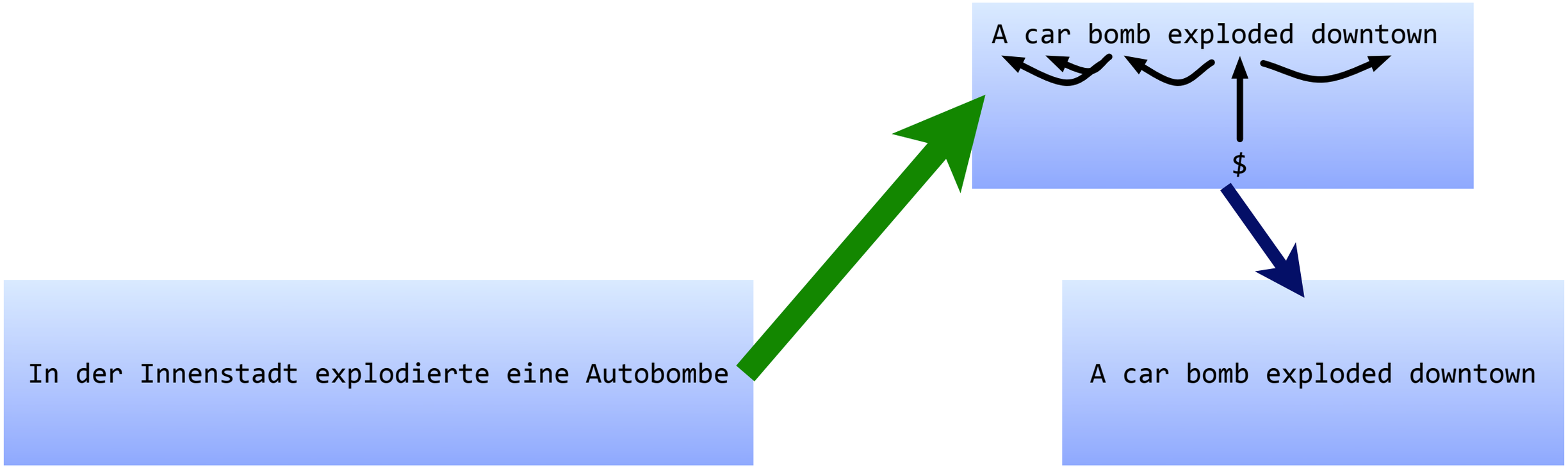

A car bomb exploded downtown
$
In der Innenstadt explodierte eine Autobombe
A car bomb exploded downtown

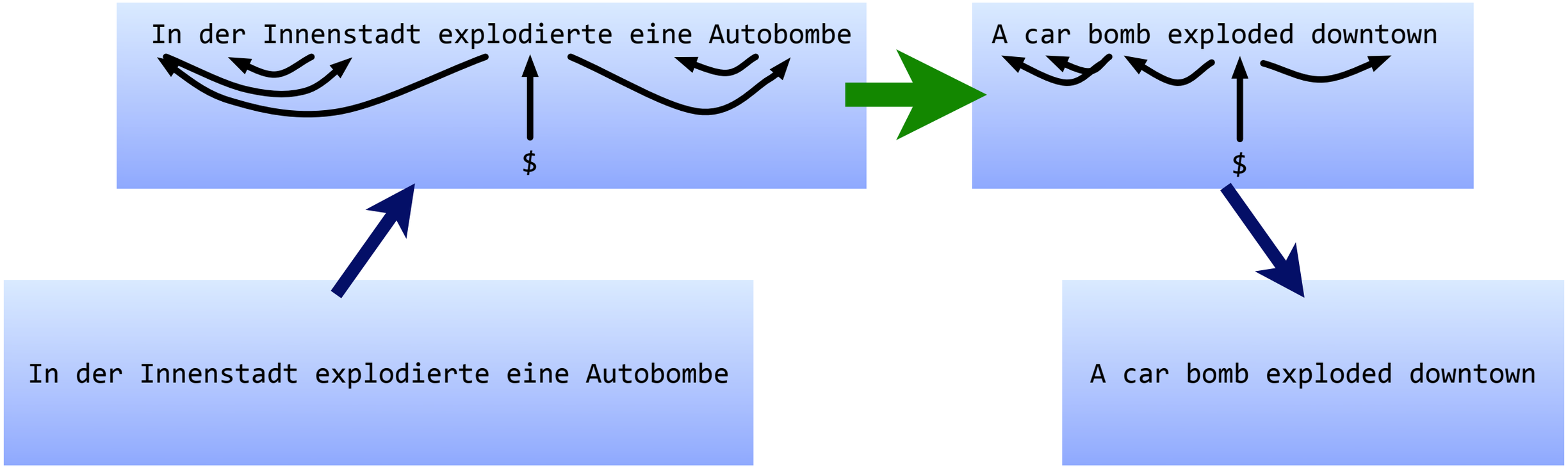
- **String-to-tree translation**
  - Transfer from source tree to target string
  - Formally a generalization of **monolingual parsing**
  - Intuition: it is more important to know the language you are translating into well than the source language



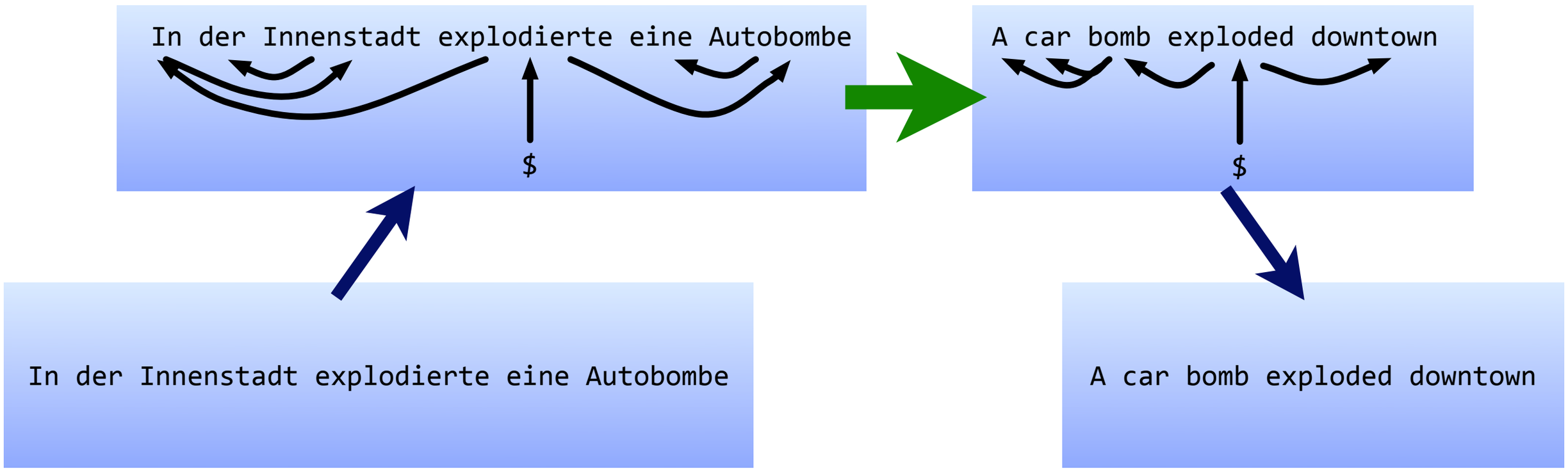
- **String-to-tree translation**
  - Transfer from source tree to target string
  - Formally a generalization of **monolingual parsing**
  - Intuition: it is more important to know the language you are translating into well than the source language

The best Chinese-English systems are string-to-tree




In der Innenstadt explodierte eine Autobombe
$
A car bomb exploded downtown
$
In der Innenstadt explodierte eine Autobombe
A car bomb exploded downtown

- **Tree to tree translation**
  - Use syntax to predict syntax
  - Benefits
    - As parsers improve, MT will improve (we hope)
    - Rich information for modeling in source and target
  - Downside: where does the syntax come from?

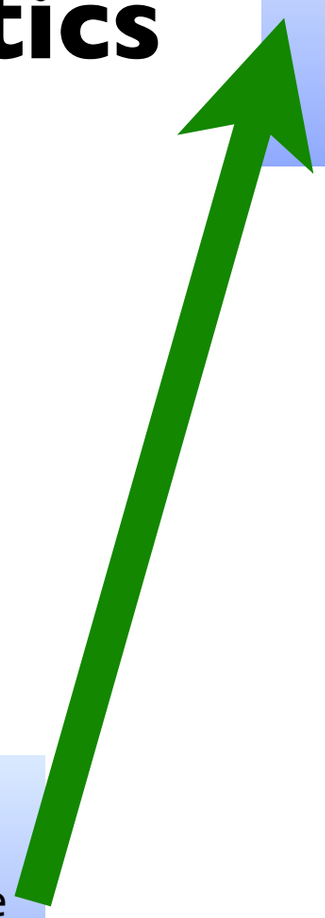


In der Innenstadt explodierte eine Autobombe

Semantics

"logical form"

```
detonate
     :arg0 bomb
     :arg1 car
     :loc downtown
     :time past
```



In der Innenstadt explodierte eine Autobombe

Semantics

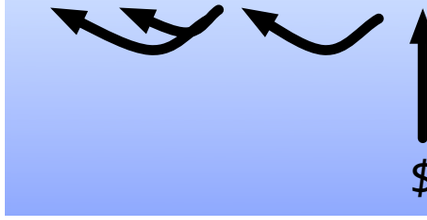
"logical form"

```
detonate
     :arg0 bomb
     :arg1 car
     :loc downtown
     :time past
```

Syntax

```
In der Innenstadt explodierte eine Autobombe
```

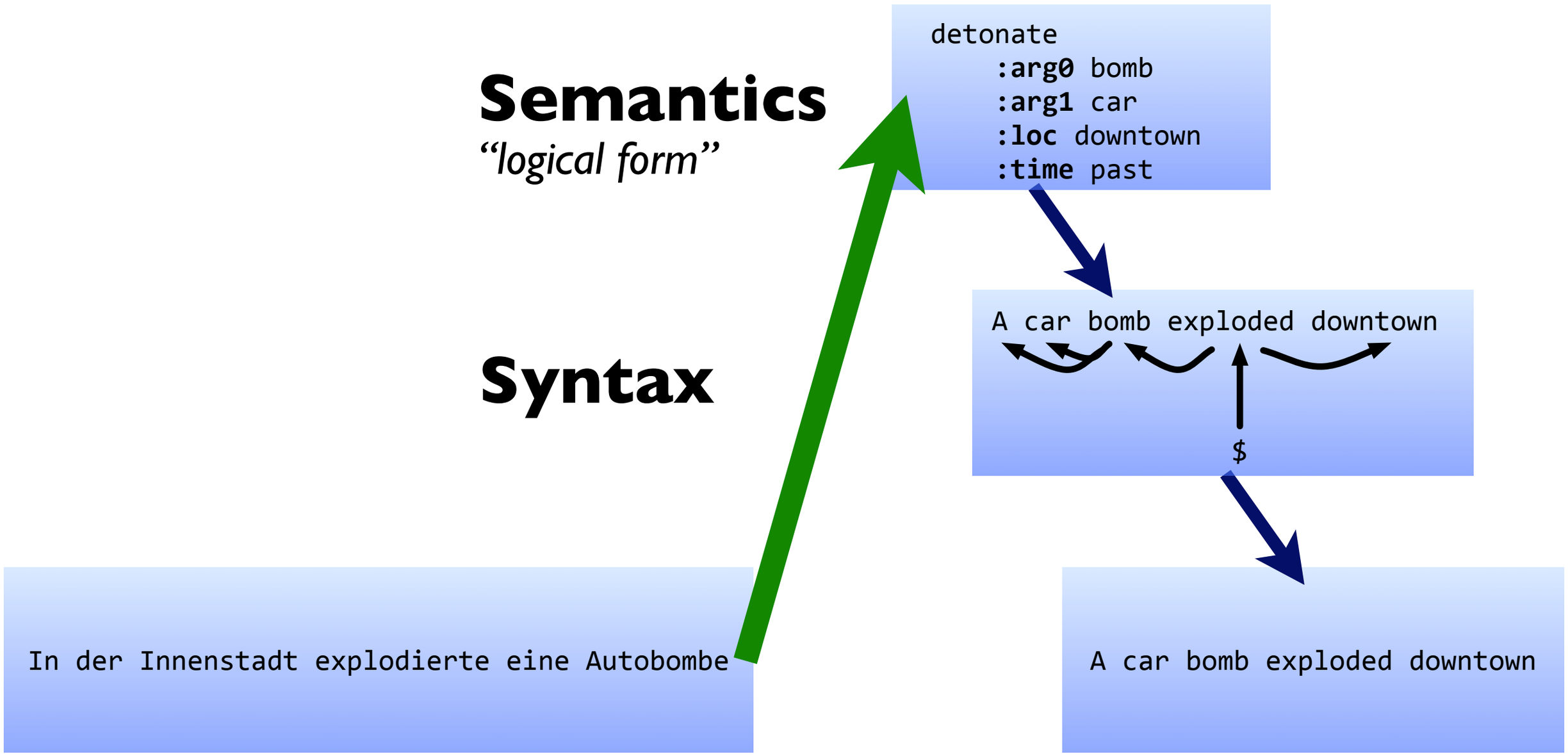
```
A car bomb exploded downtown
```

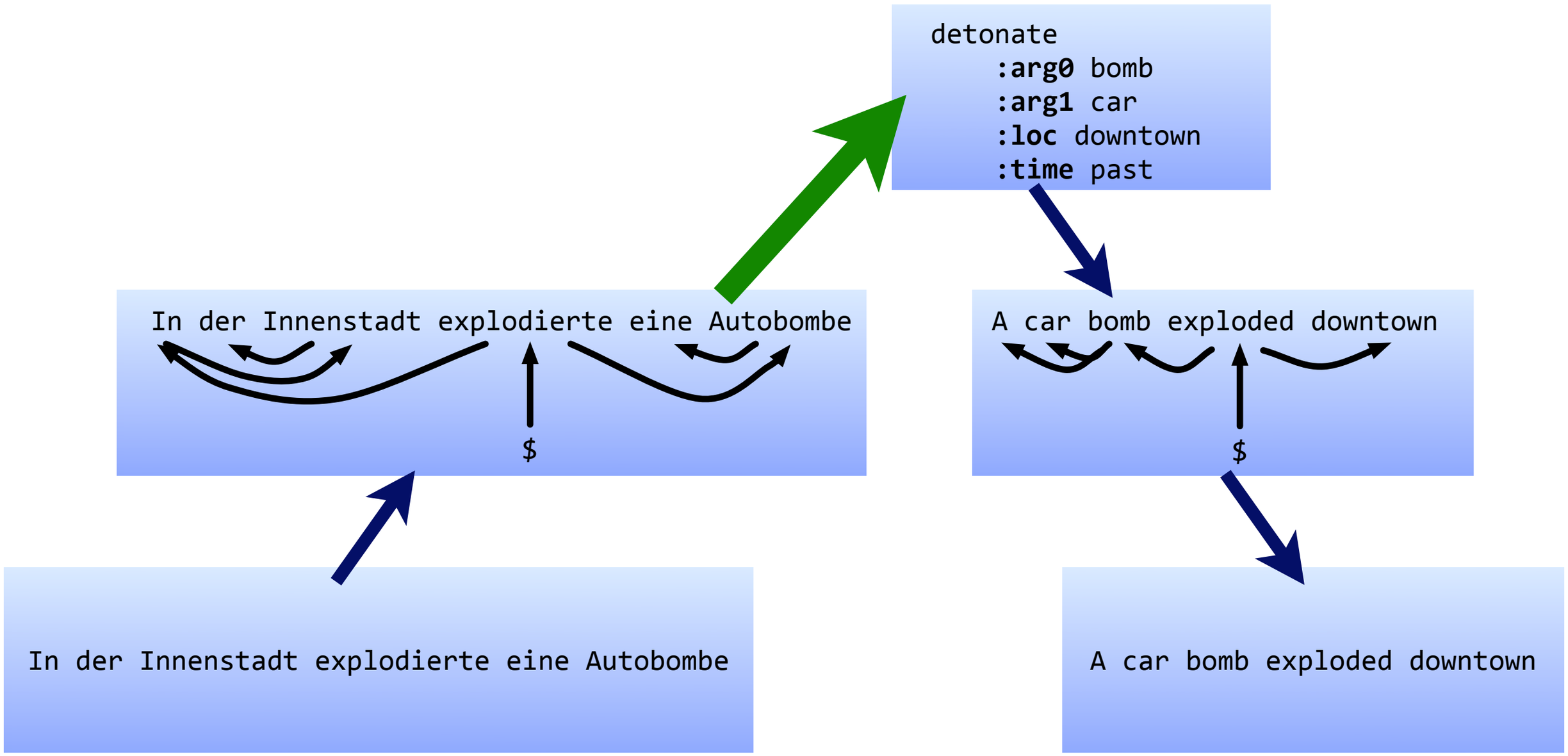


Semantics

"logical form"

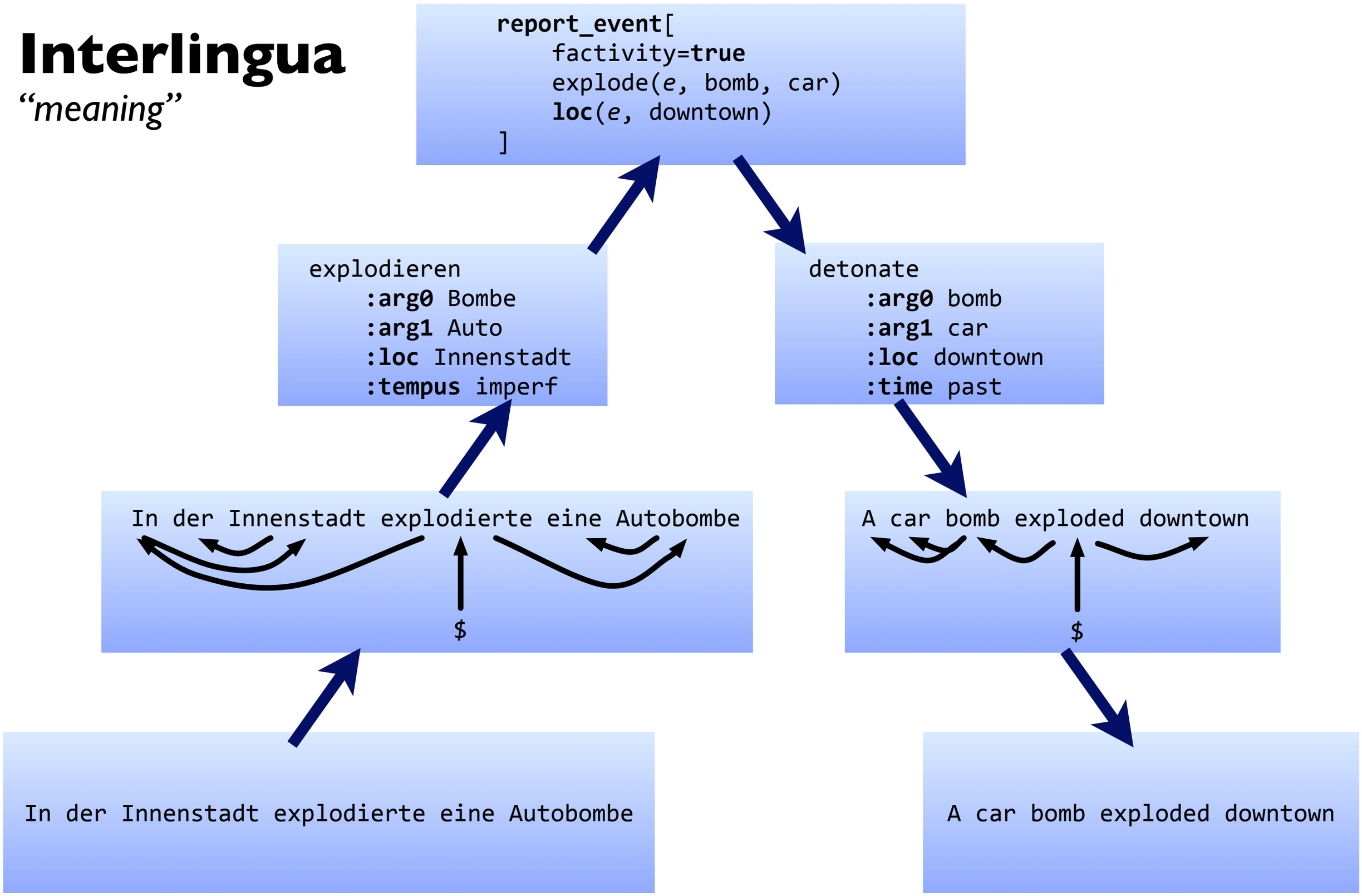
Syntax




detonate
:arg0 bomb
:arg1 car
:loc downtown
:time past
In der Innenstadt explodierte eine Autobombe
$
A car bomb exploded downtown
$
In der Innenstadt explodierte eine Autobombe
A car bomb exploded downtown

Interlingua

"meaning"



More Abstract Models

More Abstract Models

- **Modeling challenges**
  - What are the right abstract representations?
  - How do we support more abstraction without sacrificing accuracy on frequent elements?

More Abstract Models

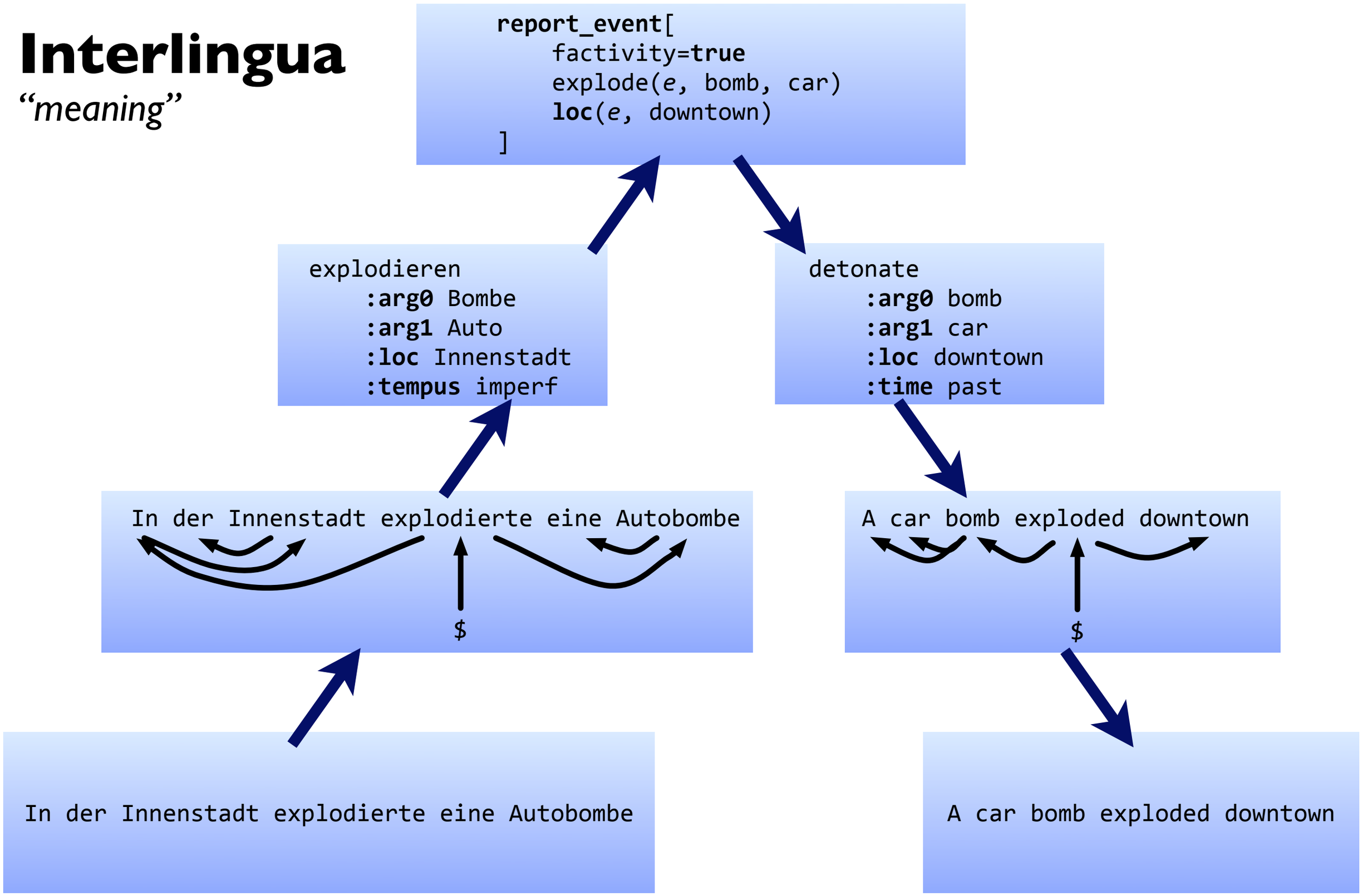
- **Modeling challenges**
  - What are the right abstract representations?
  - How do we support more abstraction without sacrificing accuracy on frequent elements?
- **Computational challenges**
  - Large search spaces
  - Error propagation in pipelines

More Abstract Models

- **Modeling challenges**
  - What are the right abstract representations?
  - How do we support more abstraction without sacrificing accuracy on frequent elements?
- **Computational challenges**
  - Large search spaces
  - Error propagation in pipelines
- **Learning challenges**
  - Nonconvexity
  - Where does the data come from?

Interlingua

"meaning"



Interlingua

"meaning"

```
report_event[
    factivity=true
    explode(e, bomb, car)
    loc(e, downtown)
]
```

```
explodieren
    arg0 Bombe
    loc Innenstadt
    tempus imperf
```

```
detonate
    :arg0 bomb
    :arg car
    :loc downtown
    :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

Other Challenges

European parliament language (training):

I declare resumed the session of the European Parliament adjourned on Friday 17 December 1999, and I would like once again to wish you a happy new year in the hope that you enjoyed a pleasant festive period.

Other Challenges

European parliament language (training):

I declare resumed the session of the European Parliament adjourned on Friday 17 December 1999, and I would like once again to wish you a happy new year in the hope that you enjoyed a pleasant festive period.

Human language (testing):

Its Me Edith :) @Lovelyyedi 40m

Birthday is on sunday & im pribably not doin anythin cause mom planned that other thing for sat nd sunday idk tf ima do -.-

Expand Reply Retweet Favorite

Other Challenges

European parliament language (training):

I declare resumed the session of the European Parliament adjourned on Friday 17 December 1999, and I would like once again to wish you a happy new year in the hope that you enjoyed a pleasant festive period.

Human language (testing):

Its Me Edith :) @Lovelyyedi 40m

Birthday is on sunday & im pribably not doin anythin cause mom planned that other thing for sat nd sunday idk tf ima do -.-

Expand Reply Retweet Favorite

spelling error

Other Challenges

European parliament language (training):

I declare resumed the session of the European Parliament adjourned on Friday 17 December 1999, and I would like once again to wish you a happy new year in the hope that you enjoyed a pleasant festive period.

Human language (testing):

Other Challenges

European parliament language (training):

I declare resumed the session of the European Parliament adjourned on Friday 17 December 1999, and I would like once again to wish you a happy new year in the hope that you enjoyed a pleasant festive period.

Human language (testing):

Its Me Edith :) @Lovelyyedi 40m

Birthday is on sunday & im pribably not doin anythin cause mom planned that other thing for sat nd sunday idk tf ima do -.-

Expand Reply Retweet Favorite

nonstandard contractions

Questions?