FONDAZIONE
BRUNO KESSLER

Introduction to Machine Translation

Marco Turchi
Fondazione Bruno Kessler – Trento, Italy
turchi@fbk.eu

Ninth Machine Translation Marathon
Trento, September 8th-13th, 2014

HLT
HUMAN LANGUAGE
TECHNOLOGY

Thanks to Chris Dyer and Marcello Federico for sharing slides and ideas.

Outline

- Motivation
  - Why Machine Translation?
  - Do we need research in Machine Translation?
  - Why is Machine Translation so Difficult?
- Approaches to MT
- Machine Translation Evaluation

Why Machine Translation?

- Information society and production of multilingual content
  - 7 billion people - 193 countries - over 150 official languages

Why Machine Translation?

- Information society and production of multilingual content
  - 7 billion people - 193 countries - over 150 official languages

- Globalization and demand for translation services
  - > 1,000 global companies operating in at least 160 countries

Why Machine Translation?

- Information society and production of multilingual content
  - 7 billion people - 193 countries - over 150 official languages

- Globalization and demand for translation services
  - > 1,000 global companies operating in at least 160 countries

- Size of worldwide translation market
  - 12.5 billion $ per year ≈ 34 million $ per day

Why Machine Translation?

- Size of translation industry
  - > 3,000 translation companies
  - > 250,000 translators

Why Machine Translation?

- Size of translation industry
  - > 3,000 translation companies
  - > 250,000 translators

- MT can improve productivity of human translators
  - integration of MT with human translation (post-editing)

Why Machine Translation?

- Size of translation industry
  - > 3,000 translation companies
  - > 250,000 translators

- MT can improve productivity of human translators
  - integration of MT with human translation (post-editing)

- MT can supply cheap gist translation
  - competitive quality-cost-speed trade-off

- …

Why Machine Translation?

- Size of translation industry
  - > 3,000 translation companies
  - > 250,000 translators

- MT can improve productivity of human translators
  - integration of MT with human translation (post-editing)

- MT can supply cheap gist translation
  - competitive quality-cost-speed trade-off

- …

Source: Common Sense Advisory, **2010**

Do we need research in Machine Translation?

Do we need research in Machine Translation?

Do we need research in Machine Translation?

Do we need research in Machine Translation?

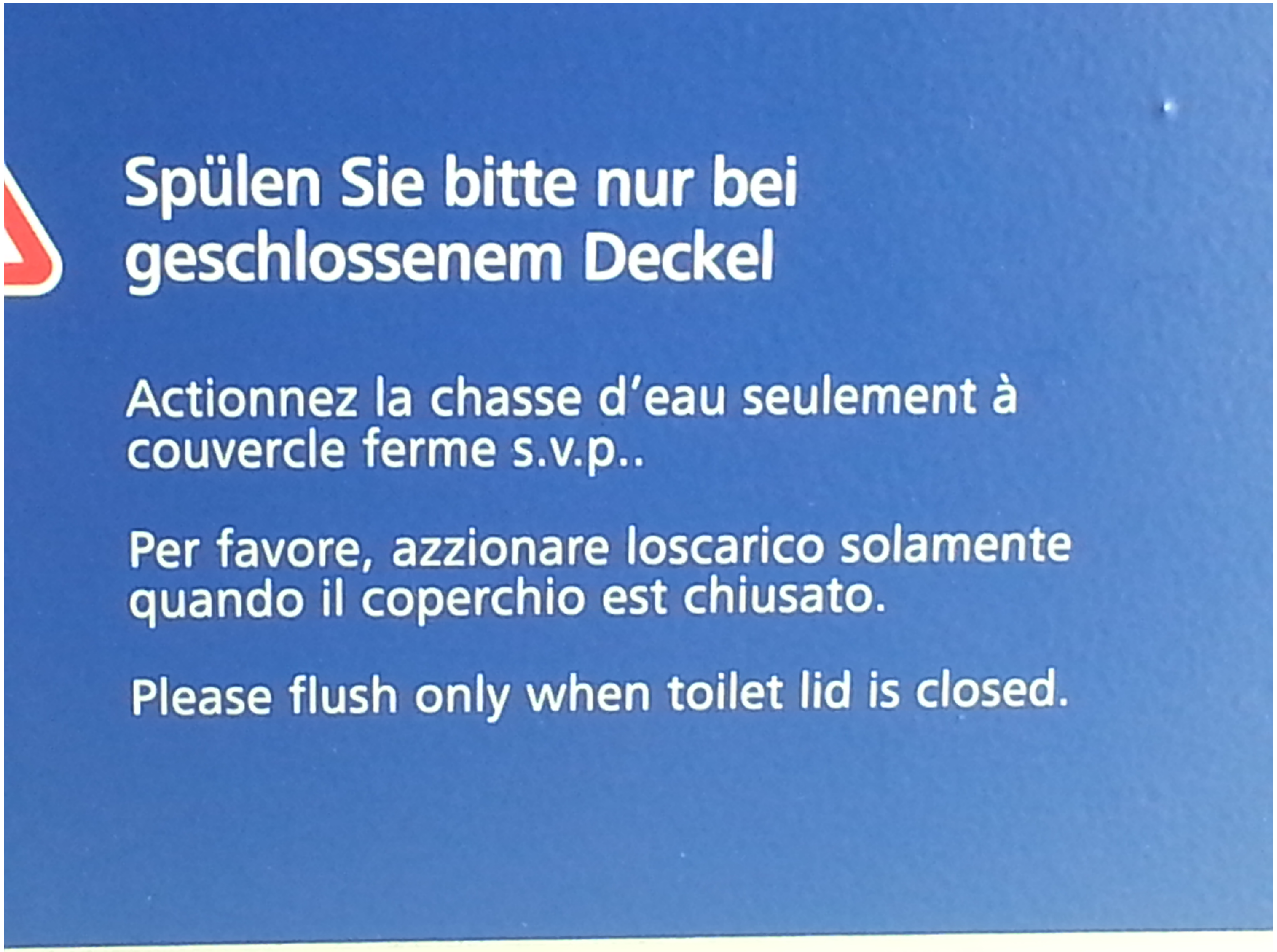
Do we need research in Machine Translation?

Chinglish examples, some of which resulting from MT errors.

Do we need research in Machine Translation?

- Chinese – English
  - difficult translation direction
  - different alphabet
  - several Chinese dialects
  - less resources than other directions
  - …

Do we need research in Machine Translation?



Why is Machine Translation so Difficult?

Why is Machine Translation so Difficult?

- High quality human translation implies:
  - deep and rich understanding of source language and text
  - sophisticated and creative command of target language

Why is Machine Translation so Difficult?

- High quality human translation implies:
  - deep and rich understanding of source language and text
  - sophisticated and creative command of target language

- Feasible goals for machine translation are tasks where:
  - even approximate translation are helpful (gist translation)
  - professional translators can take advantage of it (computer assisted translation)
  - linguistic domain is very focused and limited (apps for travelers)

Why is Machine Translation so Difficult?

- High quality human translation implies:
  - deep and rich understanding of source language and text
  - sophisticated and creative command of target language

- Feasible goals for machine translation are tasks were:
  - even approximate translation are helpful (gist translation)
  - professional translators can take advantage of it (computer assisted translation)
  - linguistic domain is very focused and limited (apps for travelers)

- Difficulty of translating depends on how similar the target and source languages are in their vocabulary, grammar, and conceptual structure.

Some Applications

Stora problem efter helgens regn

Articles : 17 | Last update : Sep 1, 2014 9:53:00 AM | Start : Sep 1, 2014 6:54:00 AM | Sources : 11 | Peak : 1 | Current rank : 1

Stora problem efter helgens regn

aftonbladet Monday, September 1, 2014 9:53:00 AM CEST | info EN [en] [other]

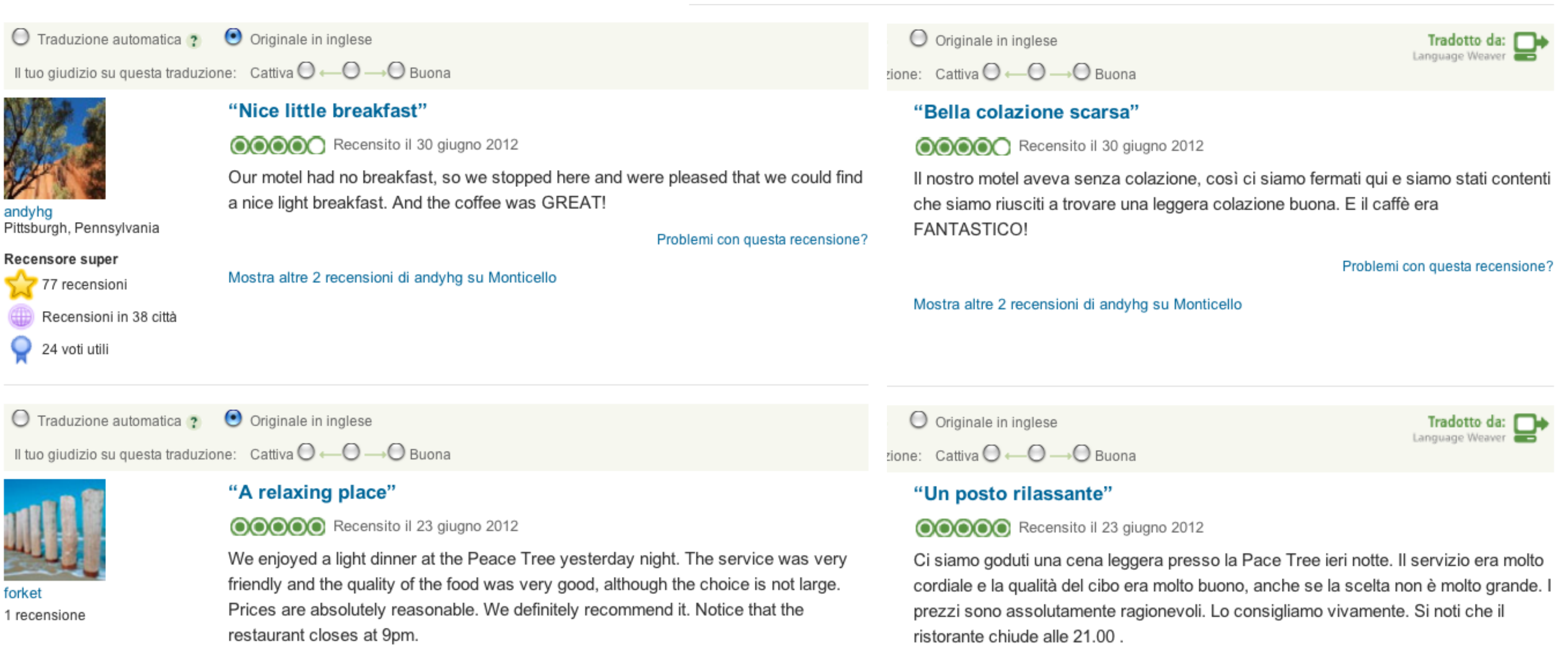
Malmö. Det råder begränsad framkomlighet på flera skånska vägar efter det kraftiga regnovädret i helgen. Även tågtrafiken påverkas i dag. Och i Malmö är hundratals hushåll utan el. Många fastigheter drabbades av stora översvämningar under gårdagen. I Malmös skyskrapa Kronprinsen har bortåt.......

> **Major problems after weekend's rain**
> Malmö. There is limited access to several skånska roads after the heavy regno weather at the weekend. Also trains are affected today. And in Malmö is hundreds of households without electricity. Many real estate was affected by major flooding yesterday. The crown prince has some Malmös skyskrapa in....

More articles...

Gist translation for social media

Some Applications

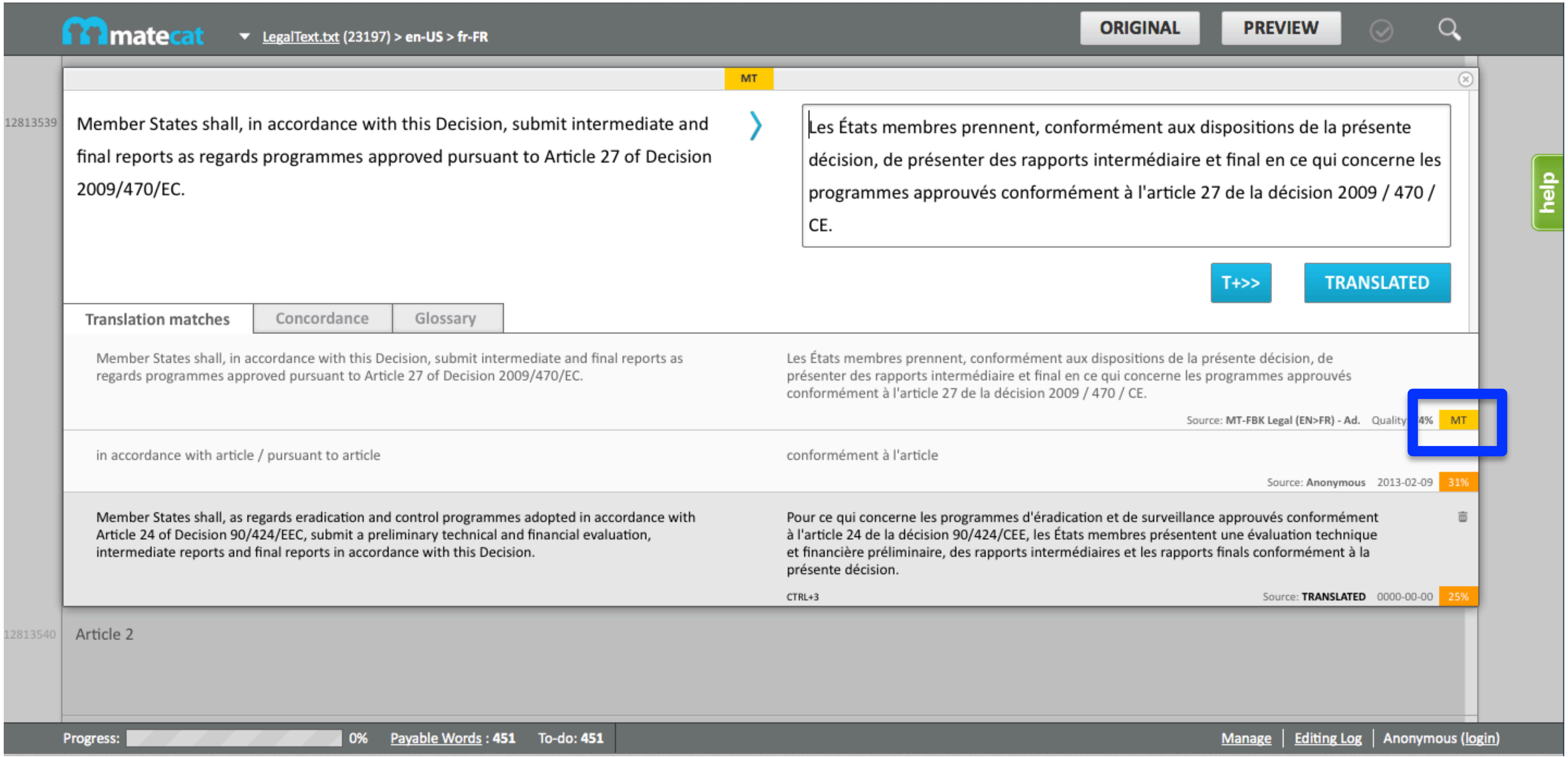


Gist translation for social media

Some Applications

Speech Translation app.

Some Applications



Integration of MT into computer assisted translation

Differences and Similarities of Languages

- Universal communicative role of language
  - names for people, words for talking about women, men, children
  - every language seems to have nouns and verbs

Differences and Similarities of Languages

- Differences/similarities across large classes of languages:
  - Morphology:
    one vs. many morphemes per words, agglutination vs. fusion
  - Syntax:
    Subj-Verb-Obj structure (E) vs. SOV (J) vs. VSO (Irish)
  - Semantics:
    mapping of semantic roles and meaning of words
    e.g. direction/manner of motion indicated by verb/satellite in the bottle floated out (E) → la botella salio flotando (S)

Differences and Similarities of Languages

- Lexical divergence between languages:
  - Semantics:
    there is no corresponding word with the same meaning
    wall (E) → Wand/Mauer (D, inside/outside)
  - Syntax:
    a word is better translated into another part-of-speech
    she likes to sing (E,v) → sie singt gerne (D,adv)

Lexical Divergence

English	brother	Japanese	otooto (younger)
			oniisan (older)
English	is	Japanese	isu (subj animate)
			aru (subj not animate)
English	know	French	connaître (be acquainted with)
			savoir (know a proposition)
English	they	French	ils (masculine)
			elles (feminine)
German	Berg	English	hill
			mountain

- Some languages make distinctions that other languages don't
- Difficulty to translate from less specific into more specific information

Outline

- Motivation
  - Why Machine Translation?
  - Do we need research in Machine Translation?
  - Why is Machine Translation so Difficult?
- Approaches to MT
- Machine Translation Evaluation

Approaches to MT

- How is knowledge and linguistic information acquired by the system?
  - Hand-crafted
  - Machine-learned

Approaches to MT

- Hand-crafted:
  - knowledge for analysis, transfer, generation, meaning representation, or direct translation is manually developed
    - most of commercial MT systems fall into this category
    - requires lots of human labor and expertise
    - includes: rule-based MT

Approaches to MT

- Machine-learned:
  – representations are implemented by mathematical models learnable from data
    - much less human effort is needed
    - requires huge amounts of data (parallel corpora of human translations), the more, the better!
    - includes: statistical MT and example-based MT

History of Machine Translation

- before 1900 - various suggestions about “mechanic” translation
- 1940s – computers used to crack the German Enigma code in World War II
- **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

Warren Weaver to Norbert Wiener, March, 1947

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***.'

Noisy Channel Model

Claude Shannon. “A Mathematical Theory of Communication” 1948.

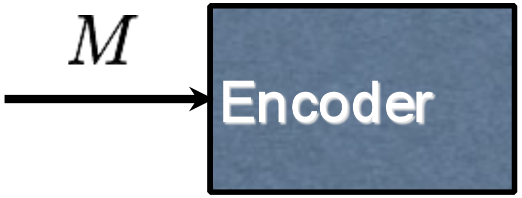
Noisy Channel Model

$$\xrightarrow{M}$$

Message

Claude Shannon. “A Mathematical Theory of Communication” 1948.

Noisy Channel Model



Message

Claude Shannon. “A Mathematical Theory of Communication” 1948.

Noisy Channel Model

Message

Sent transmission

$$p(y)$$

Claude Shannon. “A Mathematical Theory of Communication” 1948.

Noisy Channel Model

$$\xrightarrow{M} \boxed{\text{Encoder}} \xrightarrow{Y} \boxed{\text{"Noisy" channel}}$$

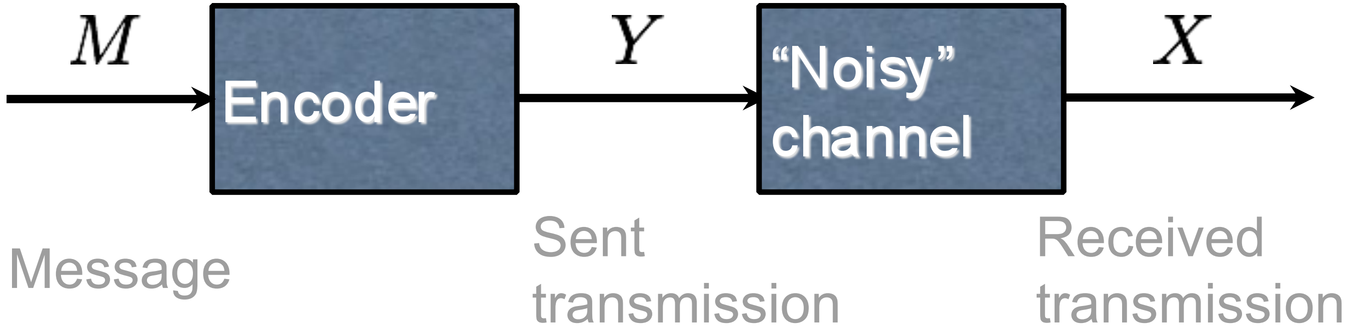
Message

Sent transmission

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

Claude Shannon. “A Mathematical Theory of Communication” 1948.

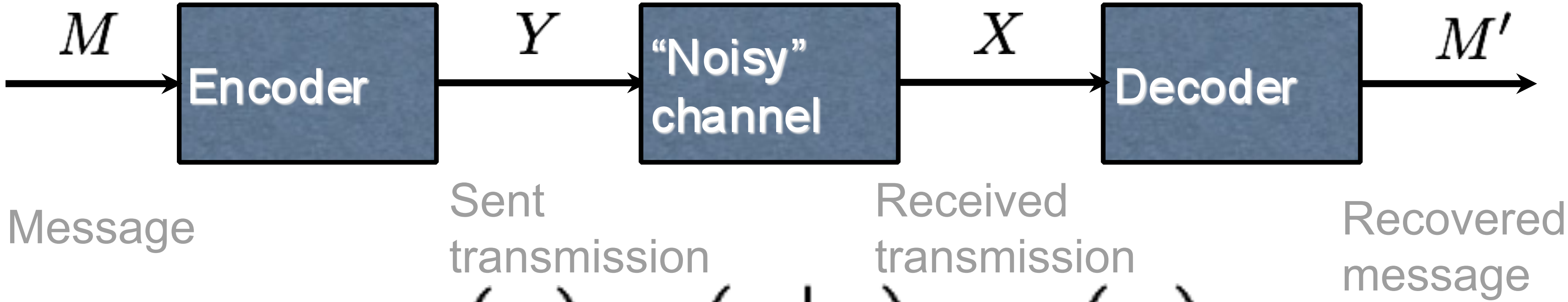
Noisy Channel Model



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

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

Noisy Channel Model



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

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

Noisy Channel Model

$$M \longrightarrow \boxed{\text{Encoder}} \xrightarrow{Y} \boxed{\text{“Noisy” channel}} \xrightarrow{X} \boxed{\text{Decoder}} \xrightarrow{M'}$$

Message | Sent | Received transmission | Recovered message

$$\boxed{p(y) \quad p(x|y)} \quad p(x)$$

Claude Shannon. “A Mathematical Theory of Communication” 1948.

Noisy Channel Model

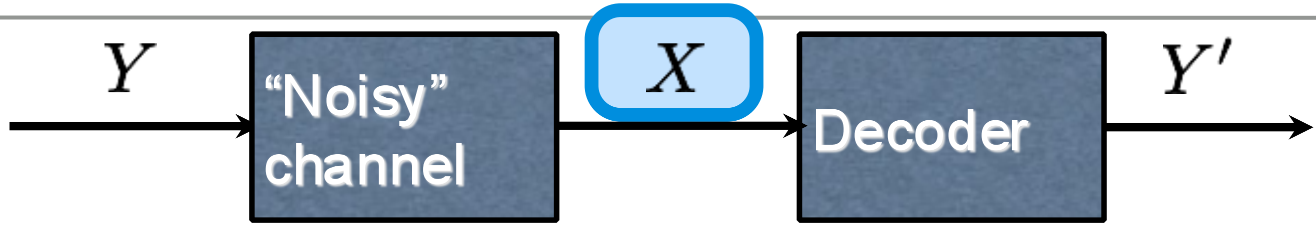
$$p(y) \quad p(x|y)$$

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.

Noisy Channel Model



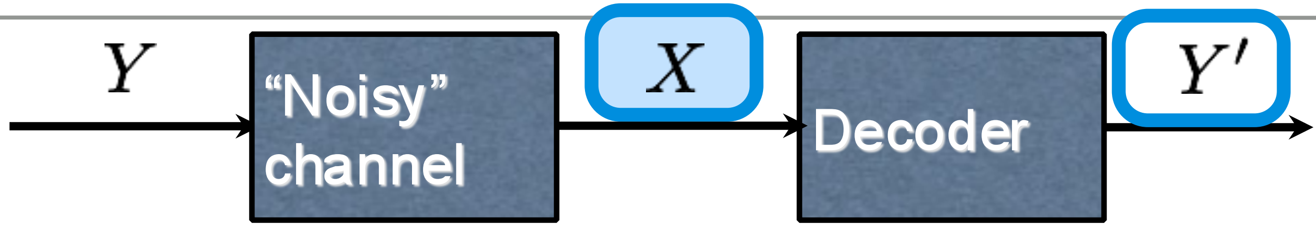
Sent transmission

Received transmission

Recovered message

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

Noisy Channel Model



Sent transmission

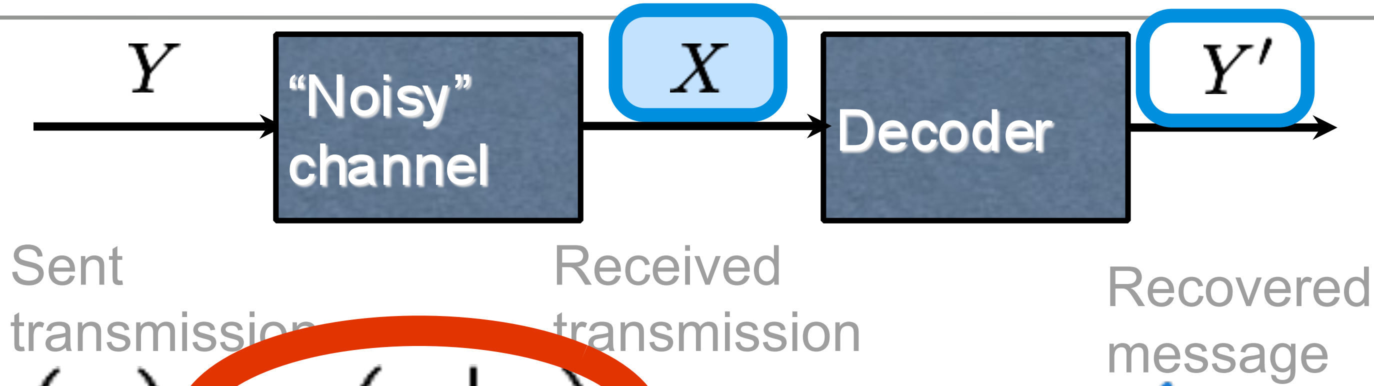
Received transmission

Recovered message

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

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

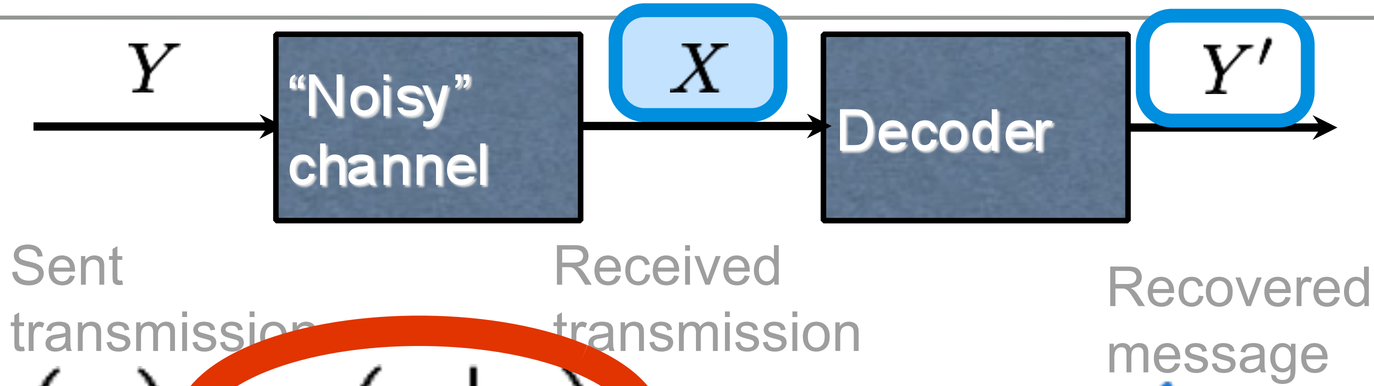
Noisy Channel Model



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

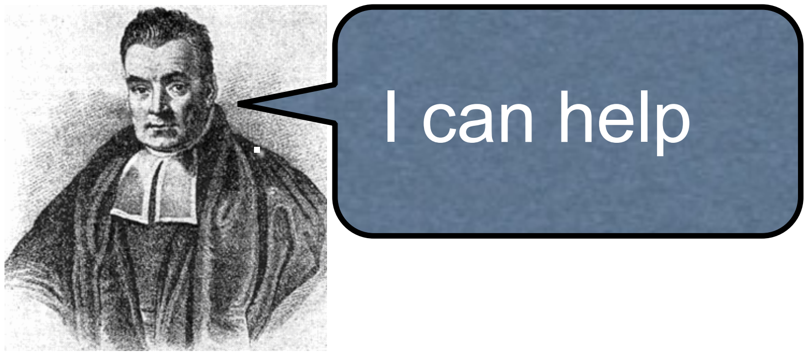
$$\boxed{y'} = \arg\max_y p(y|x)$$

Noisy Channel Model

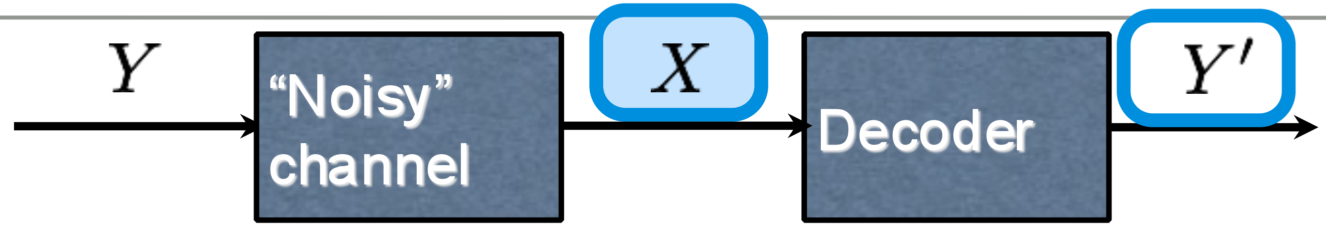


$$p(y) \quad p(x|y)$$

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



Noisy Channel Model



Sent transmission

Received transmission

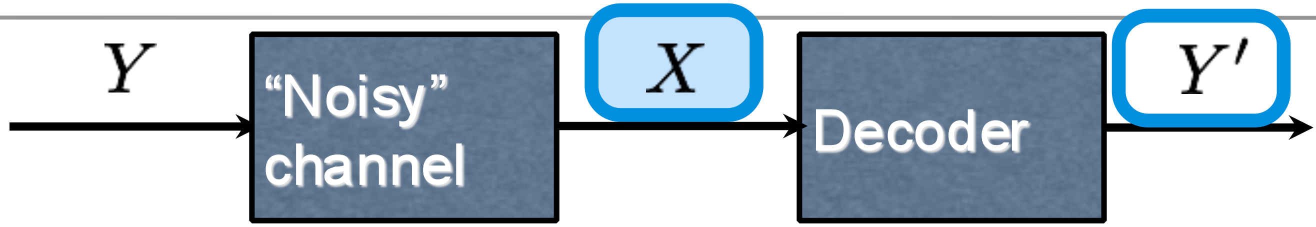
Recovered message

$p(y)$ $p(x|y)$

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

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

Noisy Channel Model



Sent transmission

Received transmission

Recovered message

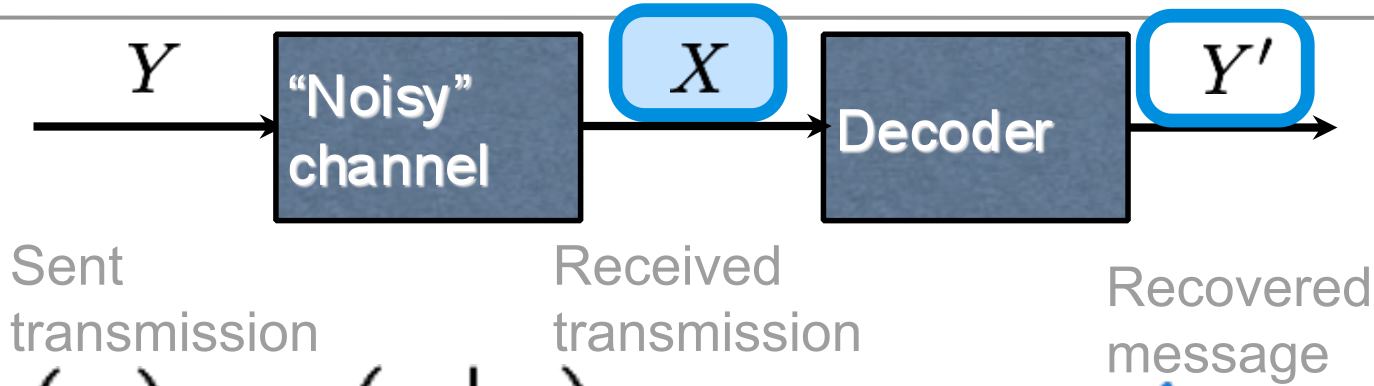
$$p(y) \qquad p(x|y)$$

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

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

Denominator does not depend on y

Noisy Channel Model



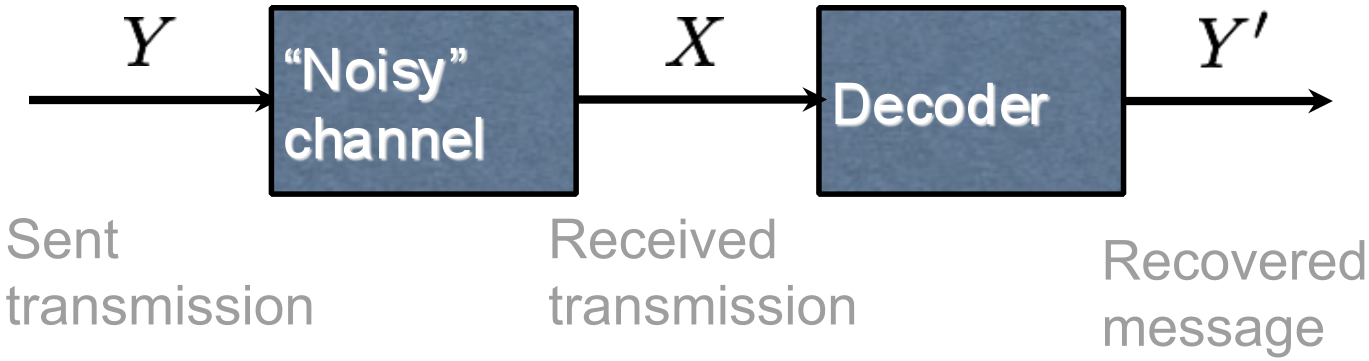
$$p(y) \qquad p(x|y)$$

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

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

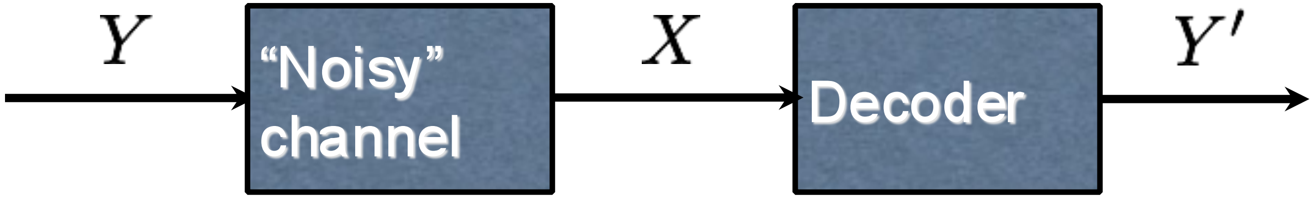
$$= \arg\max_y p(x|y)p(y)$$

Noisy Channel Model



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

Noisy Channel Model



Sent transmission

Received transmission

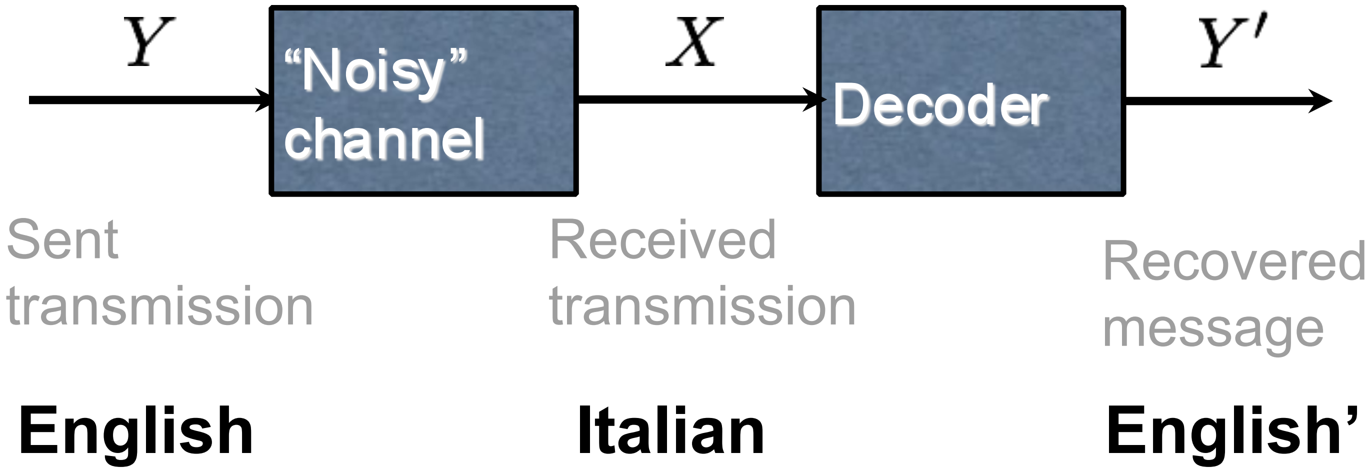
Recovered message

English **Italian** **English'**

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

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

Noisy Channel Model

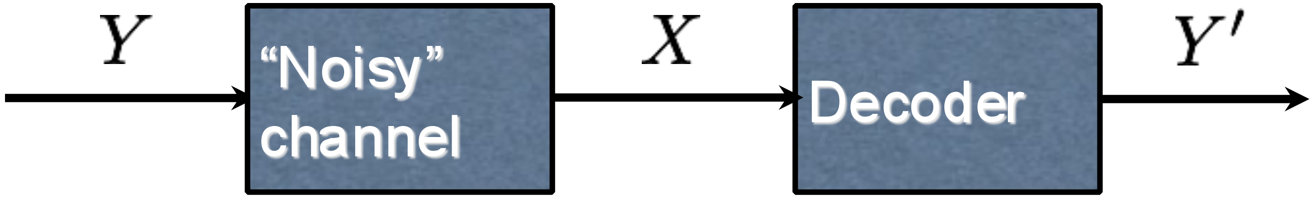


$$\cancel{y' = \arg\max_{y} p(x|y)p(y)}$$

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

translation model

Noisy Channel Model



Sent transmission | Received transmission | Recovered message

English | **Italian** | **English'**

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

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

translation model

language model

Translation Model

- provides translation *back* into the source
- learned from parallel data
  - target literally translation of the source
- guarantees **adequacy** of translation

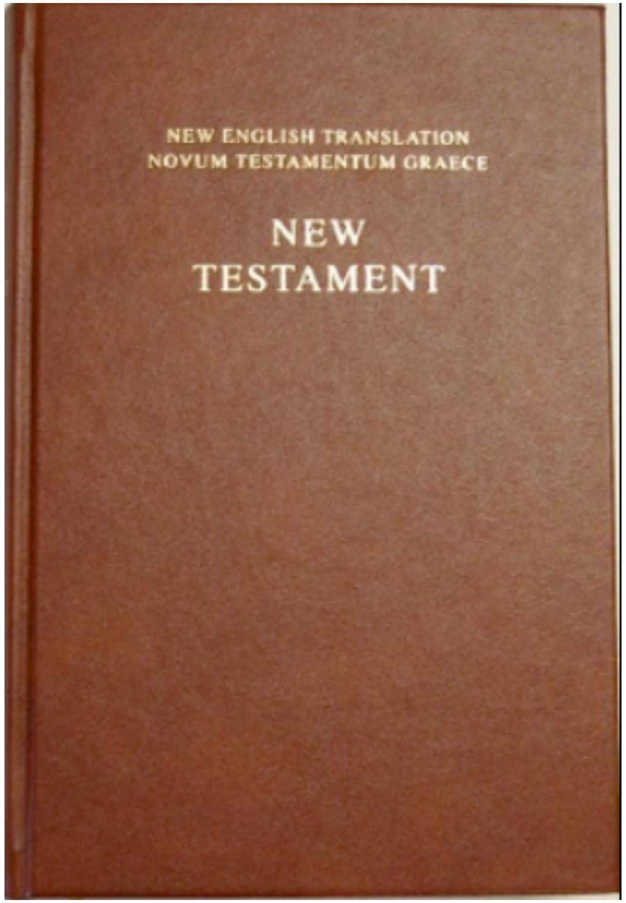
Language Model

- probability of finding a sequence of words in the target language
  - guarantees **fluency** of translation
- supports difficult decisions in word order and word translation
- learned from *any* target language corpus

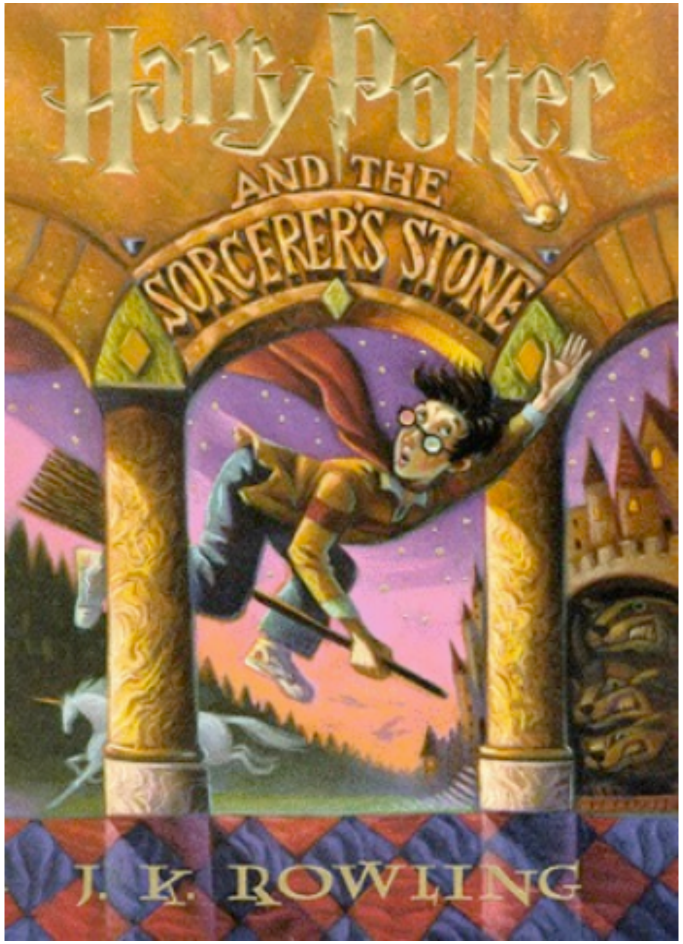
- *Tues., 9th Kenneth Heafield - Language modelling*

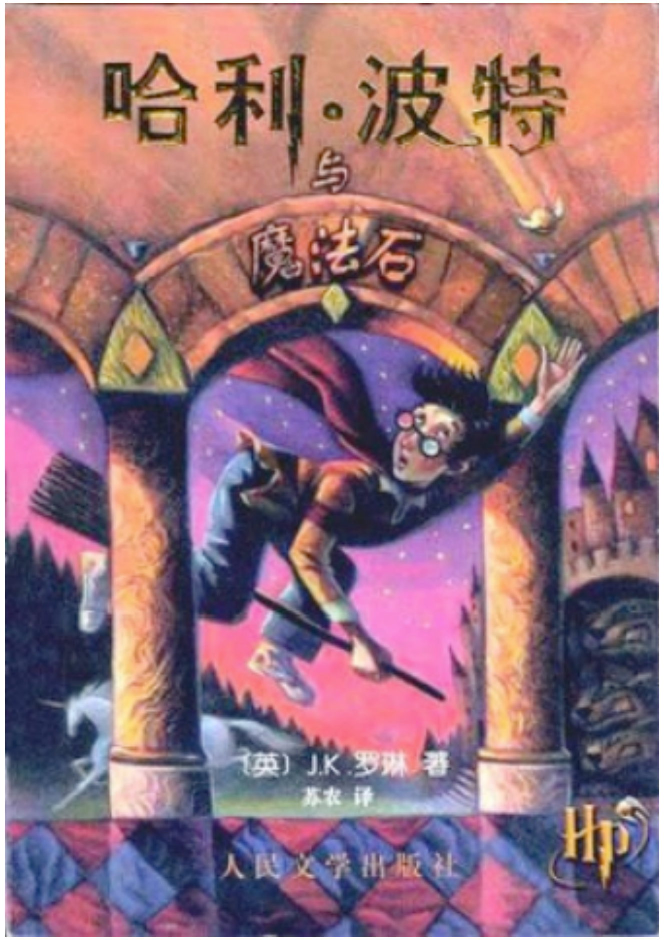
Parallel Data

Parallel Data

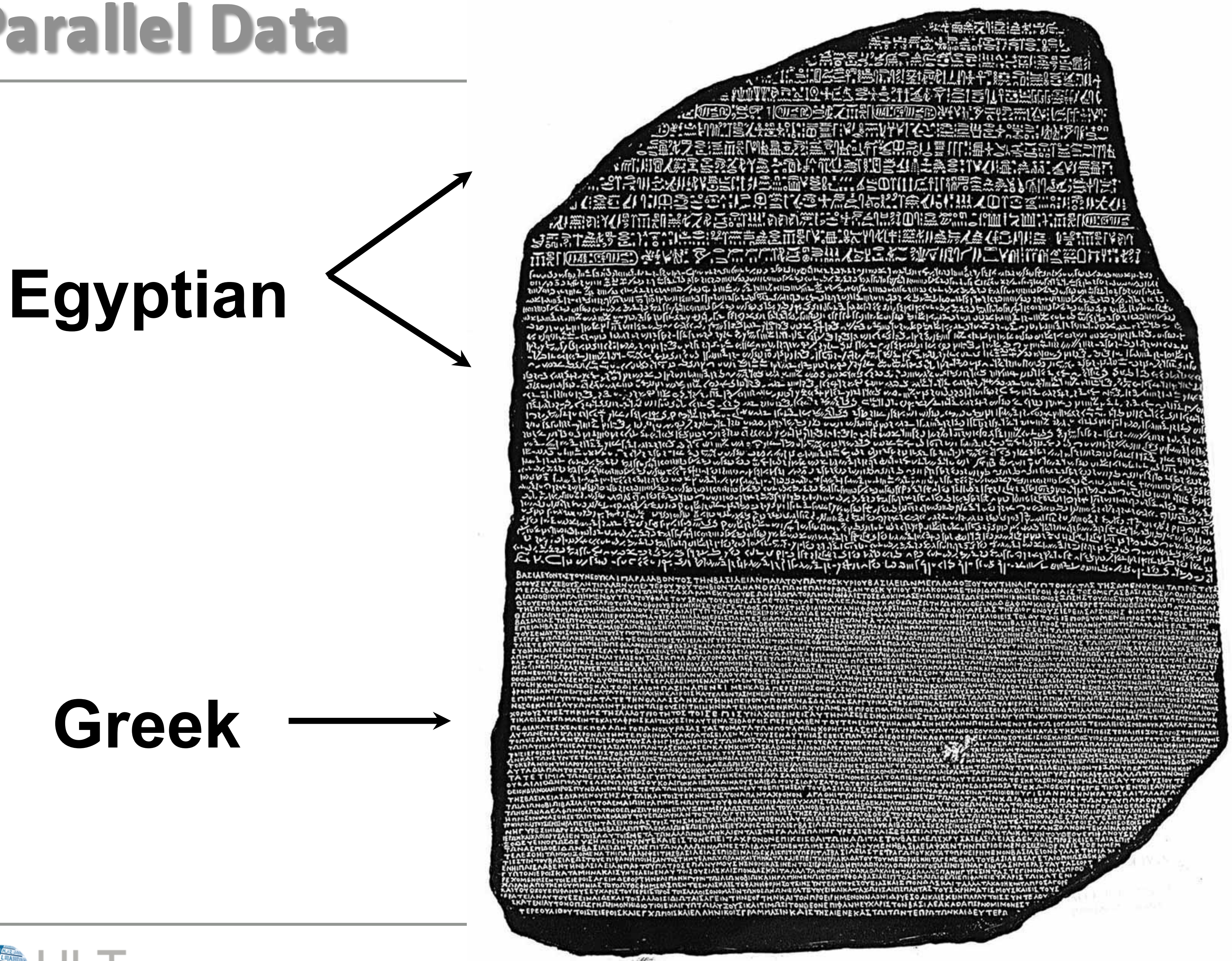


Parallel Data





Parallel Data



Statistical Machine Translation

- Parallel sentences

dalla serata di domani soffierà un freddo vento orientale

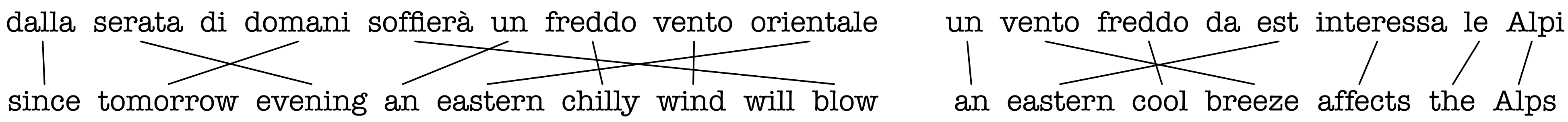
since tomorrow evening an eastern chilly wind will blow

un vento freddo da est interessa le Alpi

an eastern cool breeze affects the Alps

Statistical Machine Translation

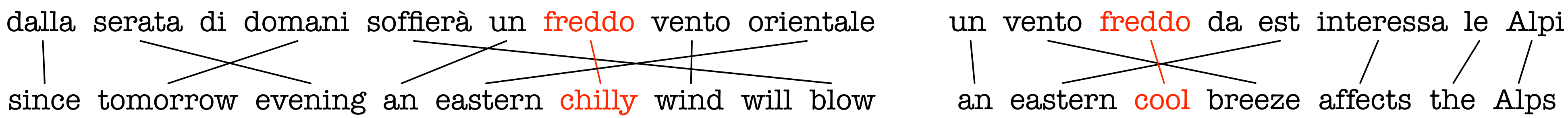
- Parallel sentences and word alignment



- Mon., 8th Mark Fishel - Word Alignment

Statistical Machine Translation

- Parallel sentences and word alignment



- Word translation probabilities

translations of **freddo**	counts	probs
chill	15	0.15
chilly	10	0.10
cold	43	0.43
cool	28	0.28
…	…	…

Statistical Machine Translation

- Parallel sentences and word alignment

dalla serata di domani soffierà un freddo vento orientale

since tomorrow evening an eastern chilly wind will blow

un vento freddo da est interessa le Alpi

an eastern cool breeze affects the Alps

- Word translation probabilities

translations of **freddo**	counts	probs
chill	15	0.15
chilly	10	0.10
cold	43	0.43
cool	28	0.28
…	…	…

translations of **vento**	counts	probs
wind	59	0.59
breeze	26	0.26
…	…	…

Statistical Machine Translation

- Parallel sentences

dalla serata di domani soffierà un freddo vento orientale

since tomorrow evening an eastern chilly wind will blow

un vento freddo da est interessa le Alpi

an eastern cool breeze affects the Alps

- Word translation probabilities

translations of **freddo**	counts	probs
chill	15	0.15
chilly	10	0.10
cold	43	0.43
cool	28	0.28
…	…	…

translations of **vento**	counts	probs
wind	59	0.59
breeze	26	0.26
…	…	…

- Word concatenation probabilities

bigrams with **eastern**	counts	probs
eastern cool	5	0.05
eastern chilly	10	0.10
eastern wind	12	0.12
eastern breeze	7	0.07
eastern …	…	…

Statistical Machine Translation

- Given word translation and concatenation probabilities

translations of **freddo**	counts	probs
chill	15	0.15
chilly	10	0.10
cold	43	0.43
cool	28	0.28
…	…	…

translations of **vento**	counts	probs
wind	59	0.59
breeze	26	0.26
…	…	…

bigrams with **eastern**	counts	probs
eastern cool	5	0.05
eastern chilly	10	0.10
eastern wind	12	0.12
eastern breeze	7	0.07
eastern …	…	…

- generate possible translations of the source sentence

un freddo vento da est

Statistical Machine Translation

- Given word translation and concatenation probabilities

translations of **freddo**	counts	probs
chill	15	0.15
chilly	10	0.10
cold	43	0.43
cool	28	0.28
…	…	…

translations of **vento**	counts	probs
wind	59	0.59
breeze	26	0.26
…	…	…

bigrams with **eastern**	counts	probs
eastern cool	5	0.05
eastern chilly	10	0.10
eastern wind	12	0.12
eastern breeze	7	0.07
eastern …	…	…

- generate possible translations of the source sentence

un freddo vento da est	a cool eastern breeze	0.08
	an eastern chilly wind	0.10
	a eastern cool wind	0.09
	a cold eastern wind	0.12
	an eastern chilly breeze	0.05
	…	…

Statistical Machine Translation

- Given word translation and concatenation probabilities

translations of **freddo**	counts	probs
chill	15	0.15
chilly	10	0.10
cold	43	0.43
cool	28	0.28
…	…	…

translations of **vento**	counts	probs
wind	59	0.59
breeze	26	0.26
…	…	…

bigrams with **eastern**	counts	probs
eastern cool	5	0.05
eastern chilly	10	0.10
eastern wind	12	0.12
eastern breeze	7	0.07
eastern …	…	…

- generate possible translations of the source sentence

un freddo vento da est	a cool eastern breeze	0.08
	an eastern chilly wind	0.10
	a eastern cool wind	0.09
	a cold eastern wind	0.12
	an eastern chilly breeze	0.05
	…	…

- return the best scoring translation

Phrase-based Model

- Phrase:
  - sequence of words without any linguistic notion
- Phrases as atomic elements
  - words not be the best atomic units, due to many to many mapping.
- Advantages:
  - translating word groups helps to resolve translation ambiguities
  - given a large training corpora, longer and longer phrases can be learnt.
- *Wed., 10^{th} Ulrich Germann - Phrase based model*

Hierarchical Phrase-based Model

- Discontinuous phrases, i.e. phrases with gaps
- Long-range reordering rules
- Formalized as synchronous context-free grammars
- no linguistic syntax, just a formally syntactic model
- The model is fully machine learnable!

Syntax-based Model

- Linguistic syntax
- Non-terminals for words and phrases: np, vp, pp, adj, ...
- Corpus annotated with syntactic parsers
- *Thurs., 11th Marcello Federico - Hierarchical and Syntactic Models*

Outline

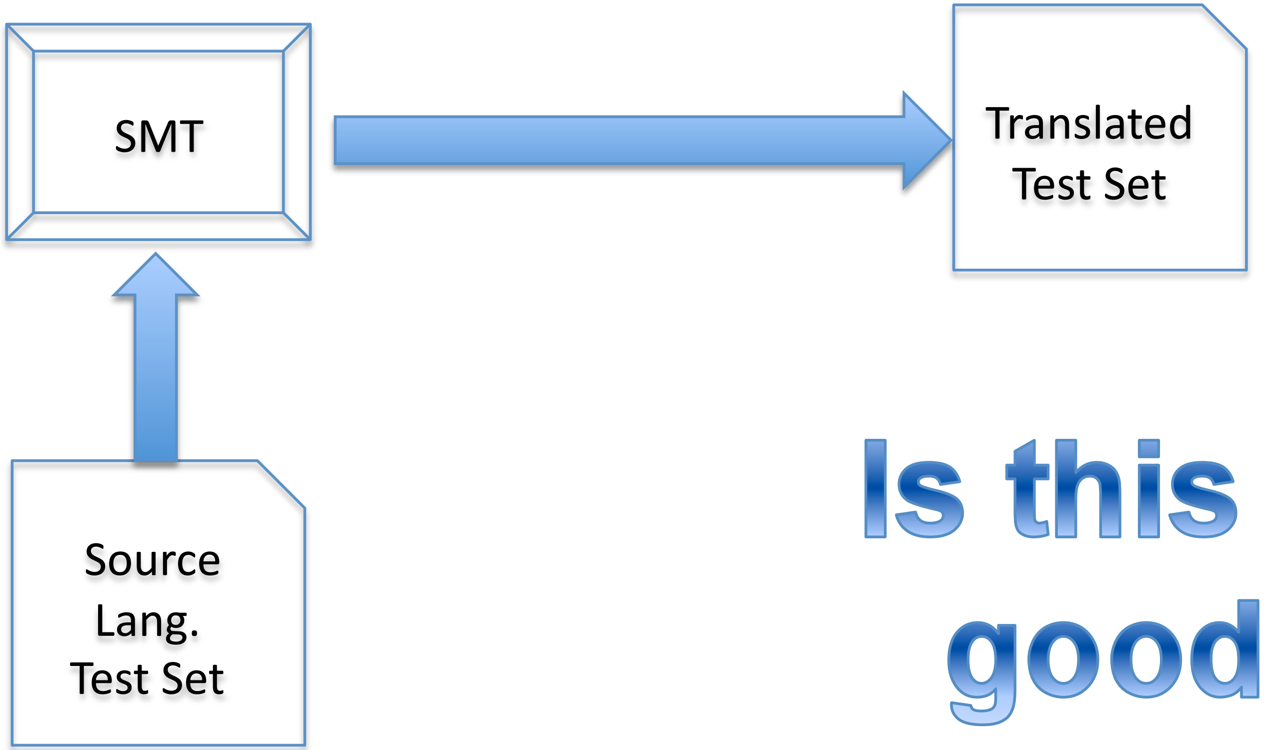
- Motivation
  - Why Machine Translation?
  - Do we need research in Machine Translation?
  - Why is Machine Translation so Difficult?

- Approaches to MT

- Machine Translation Evaluation

Machine Translation Evaluation

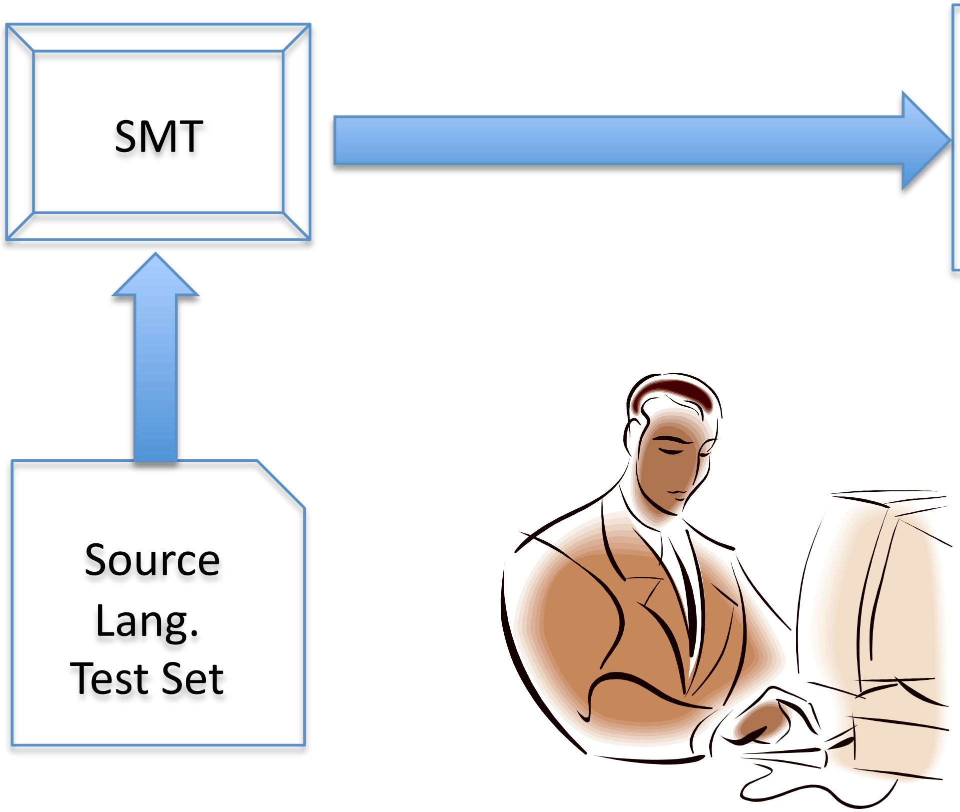
- Automatic Evaluation of MT output

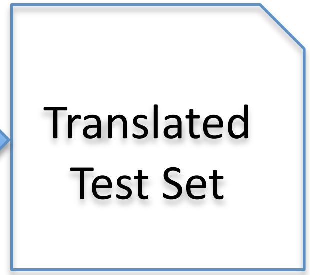


Is this a good translation?

Machine Translation Evaluation

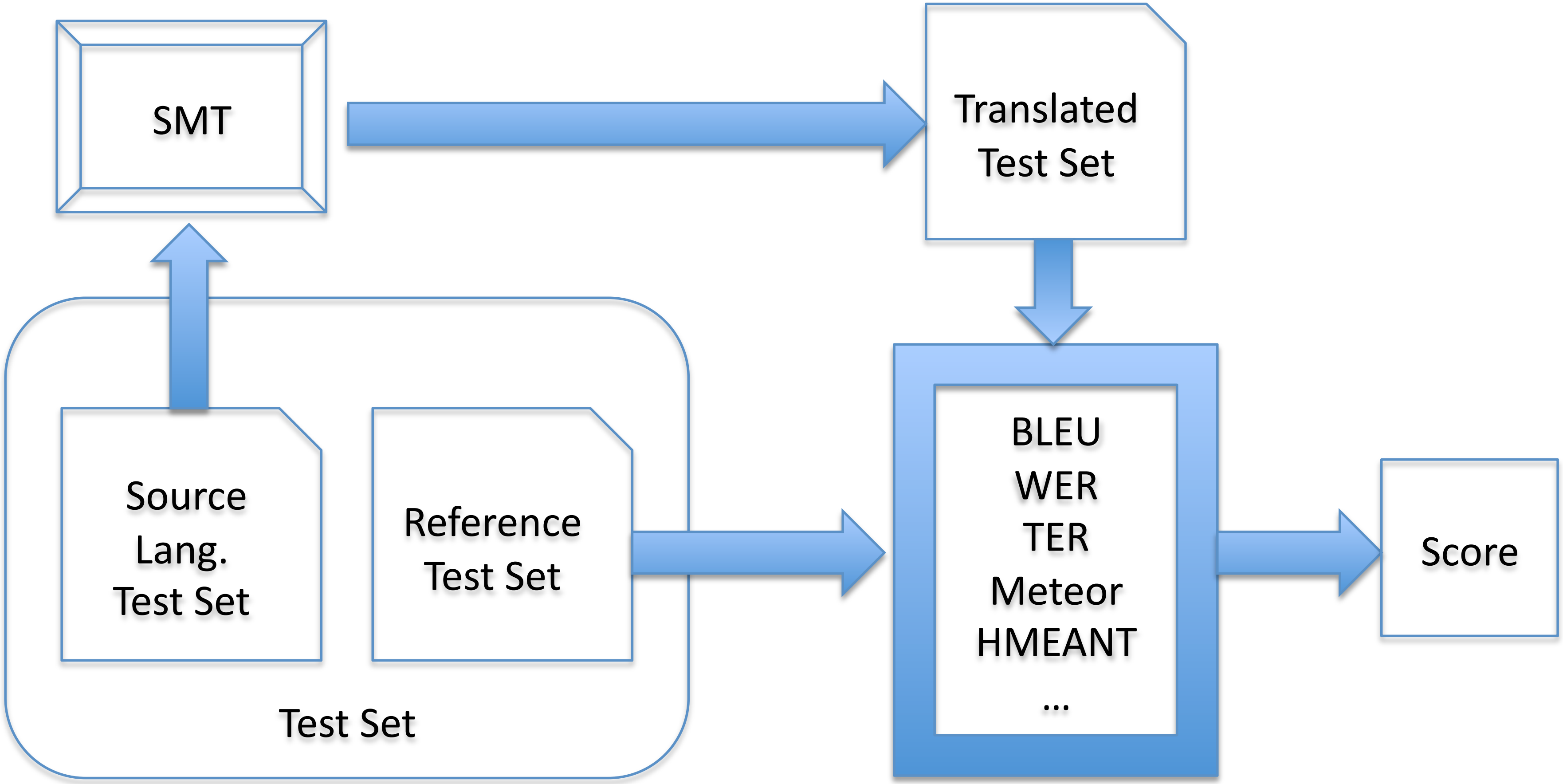
- Human presence





- … but it is:
  - subjective
  - time consuming
  - difficult
  - not replicable

Machine Translation Evaluation



- Reference Test set: human translation of the source language test set into the target language.

Machine Translation Evaluation

- Large amount of parallel test sets in different languages

- Automatic Scoring Methods (Bleu, TER, WER, Meteor):
  - low cost (wrt human evaluation)
  - objective (unbiased)
  - informative (for system developers): to profile system behavior
  - discriminative: to tell if and where improvements are
  - effective and replicable: to be computed quickly and often

- *Tues., 9th   Maja Popovic - MT evaluation,QE*

Challenges

- Translation into and from morphologically rich languages
  - Philip Williams - Morphology in SMT
- Training and test data sampled for different domains
  - Marine Carpuat - Domain adaptation in MT
- Model able to better generalize from training data
  - Holger Schwenk - Deep Learning for MT
- Translation of specific texts
  - Bruno Pouliquen - Patent translation

Challenges

- Document translation
  - Bonnie Webber - Discourse in SMT
- Translation from the crowd
  - Joao Graca - Crowdsourcing for MT
- Interaction between Human and Machine Translation System
  - Francisco Casacuberta - Interactive MT
- Post-editing Machine Translated Output
  - Sharon O'Brien - Post-editing
  - Marco Trombetti - CAT tools

Questions

FONDAZIONE
BRUNO KESSLER

Introduction to Machine Translation

Marco Turchi
Fondazione Bruno Kessler – Trento, Italy
turchi@fbk.eu

Ninth Machine Translation Marathon
Trento, September 8th-13th, 2014

HLT
HUMAN LANGUAGE TECHNOLOGY

Thanks to Chris Dyer and Marcello Federico for sharing slides and ideas.