DATA DRIVEN MACHINE TRANSLATION

David Talbot

22nd April 2017

Computer Science Club, St. Petersburg, Russia

COURSE OVERVIEW

what you will learn

- Why machine translation is hard
- How to build a machine translation system from data
- How to evaluate a machine translation system
- How to improve it using linguistic knowledge/inductive bias

what you will learn

- Phrase-based machine translation
  - Word alignment
  - Syntax based reordering
  - Language models
- Neural machine translation
  - Word embeddings
  - Encoder-decoder models
  - Attention mechanisms
  - Challenges for NMT

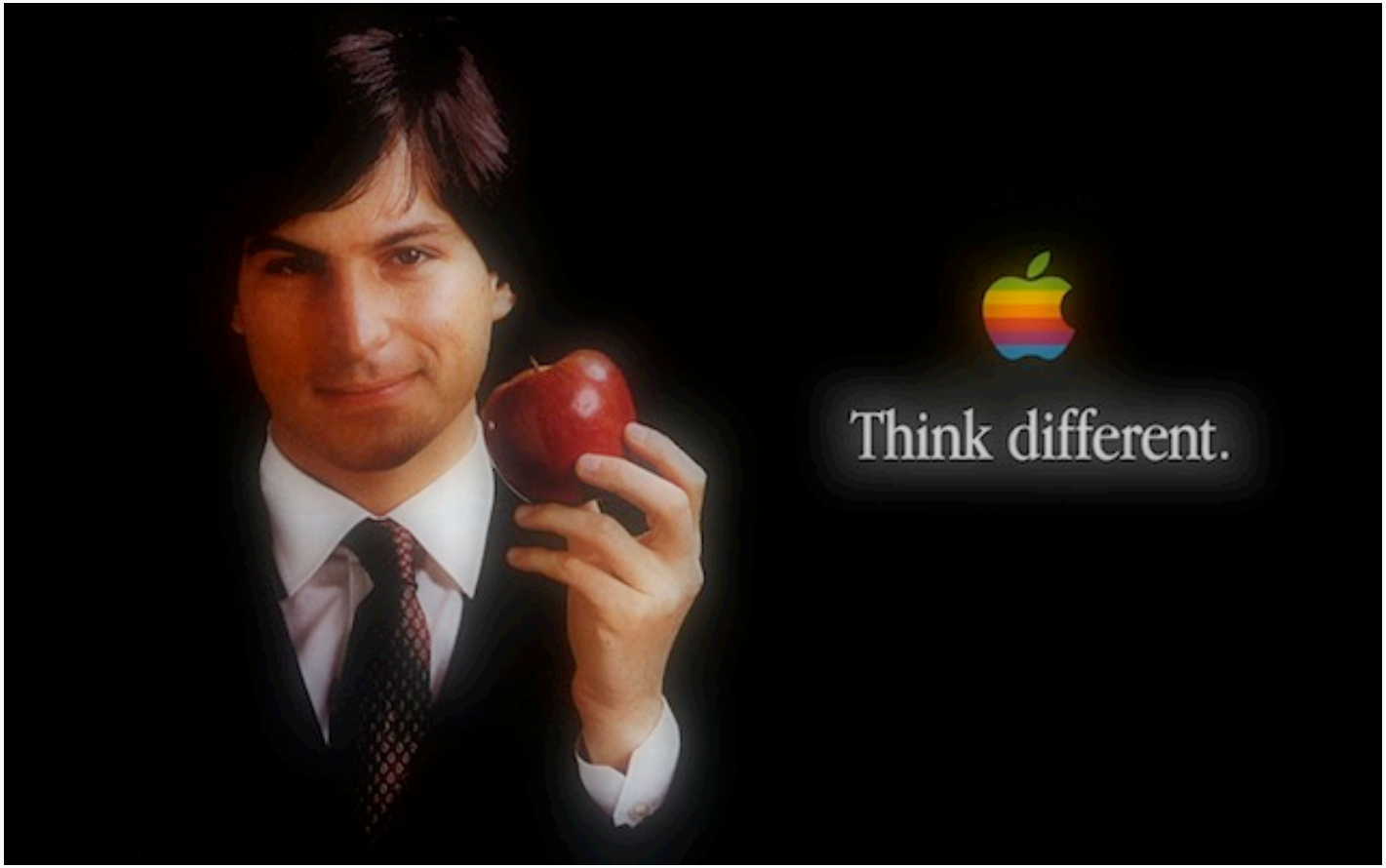
what else you'll learn (beyond mt)

Practical aspects of statistical modelling

- Estimating statistical models from data
- Handling hidden variables
- Introducing inductive bias into statistical models

WHY TRANSLATION IS HARD

ambiguity



”Finally a computer that understands you like your mother.”
(Apple, 1985)

ambiguity

”Finally a computer that understands you like your mother.”

ambiguity

”Finally a computer that understands you like your mother.”

- В конце концов компьютер, который понимает вас так же хорошо, как ваша мама (понимает вас).

ambiguity

”Finally a computer that understands you like your mother.”

- В конце концов компьютер, который понимает вас так же хорошо, как ваша мама (понимает вас).
- В конце концов компьютер, который понимает, что вам нравится ваша мама.

ambiguity

”Finally a computer that understands you like your mother.”

- В конце концов компьютер, который понимает вас так же хорошо, как ваша мама (понимает вас).
- В конце концов компьютер, который понимает, что вам нравится ваша мама.
- В конце концов компьютер, который понимает вас так же хорошо, как он понимает вашу маму.

word order

"As English not all languages words in the same order put. Hmmmmmm." – Yoda

word order

Languages can use very different word order

- He went to school by train.
- 彼は電車で学校に行きました。
- kara wa densha de gakkou ni ikimashita.

Why is reordering a huge problem for MT systems?

word order

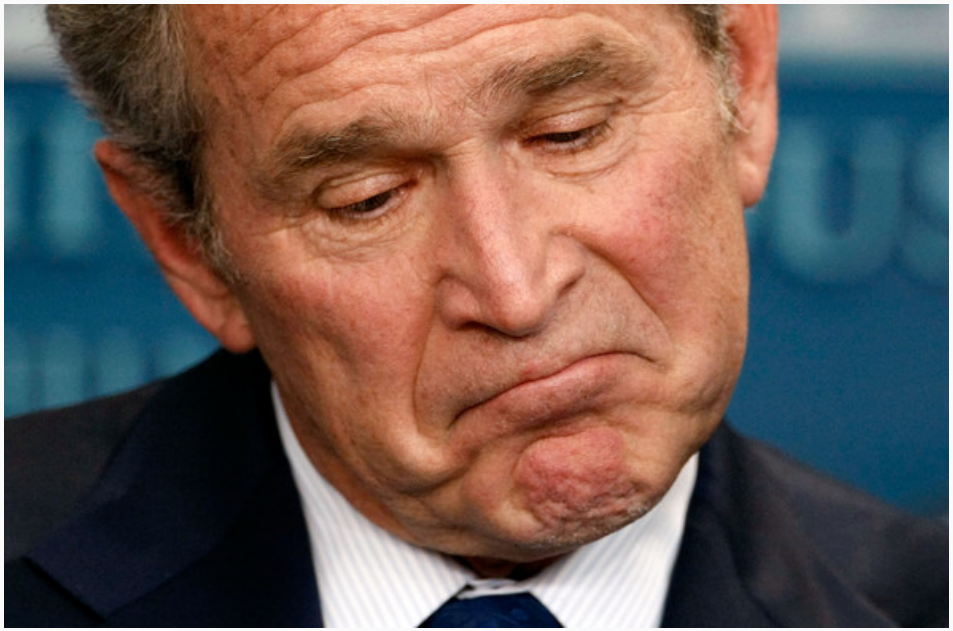
Languages can use very different word order

- He went to school by train.
- 彼は電車で学校に行きました。
- kara wa densha de gakkou ni ikimashita.

Why is reordering a problem for an MT system?

- A priori set of possible 'reorderings' scales exponentially with sentence length

morphology



“Rarely is the question asked: Is our children learning? “

– George W. Bush

morphology

Many languages require case marking and agreement

- Перечень различных рекордов скорости, установленных на рельсовых путях, был ...
- Word choices depend on the gender, case, number etc. of other words
- Morphological agreement can span many tokens

information structure

Morphological case and agreement make word order less important

- The dog bit the hippopotamus.
- The hippopotamus bit the dog.

- Собака укусила бегемота.
- Собаку укусил бегемот.

world knowledge

Usually only one interpretation is reasonable for us

- Stolen painting found by tree.
- I haven’t slept for ten days.
- I saw a man with a telescope.

even *easy* sentences are hard

en I've got two brothers.

fr J'ai deux frères.
(I've two brothers.)

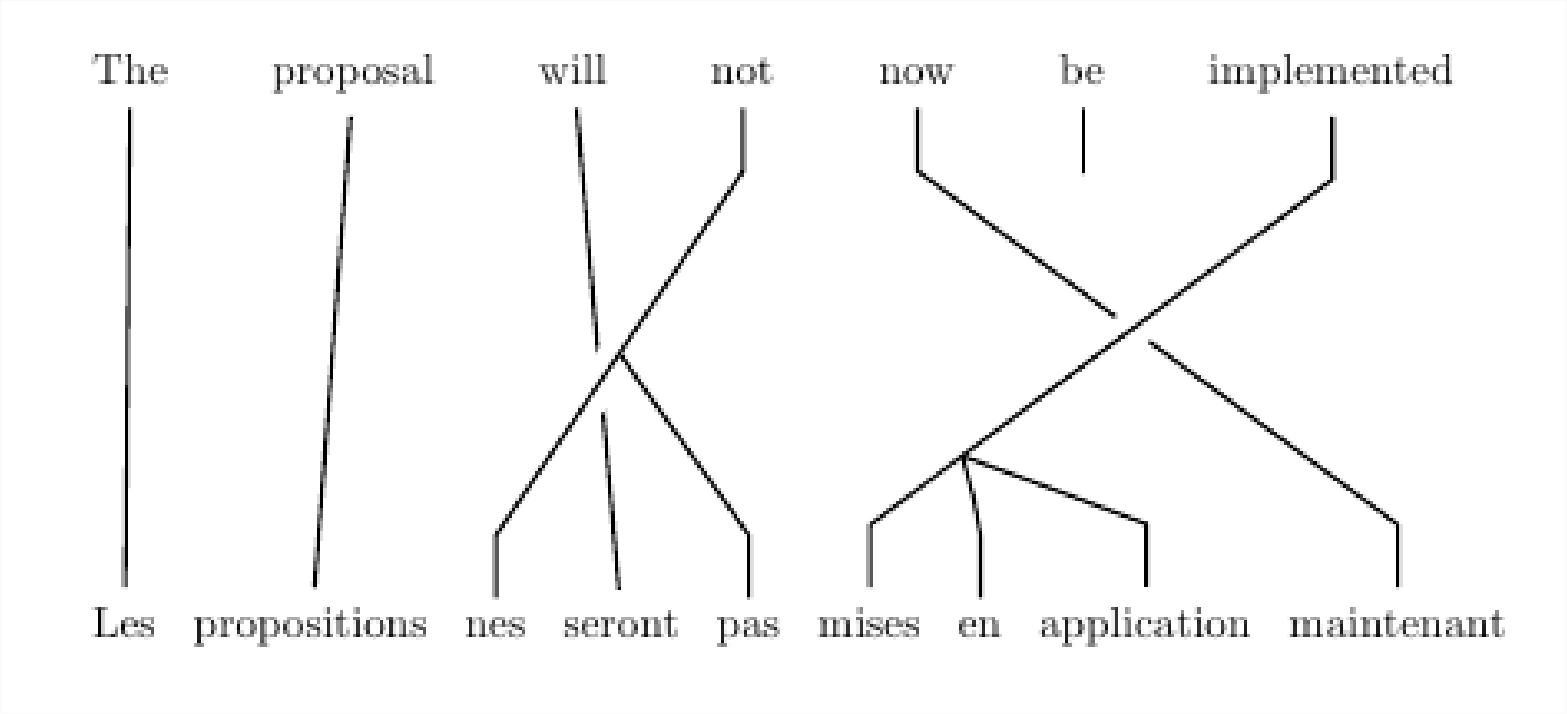
ru У меня два брата.
(At me [are] two brothers.)

ja 私は 2 人の弟がいます.
(As for me, two people younger brother there are.)

not a simple machine learning problem

- High dimensional: vocabulary > 1 million words
- Sparse: natural language follows a Zipf law
- Combinatorial: reordering is *a priori* $O(N!)$
- Dependencies
- More than one right answer...
- Partially observed data

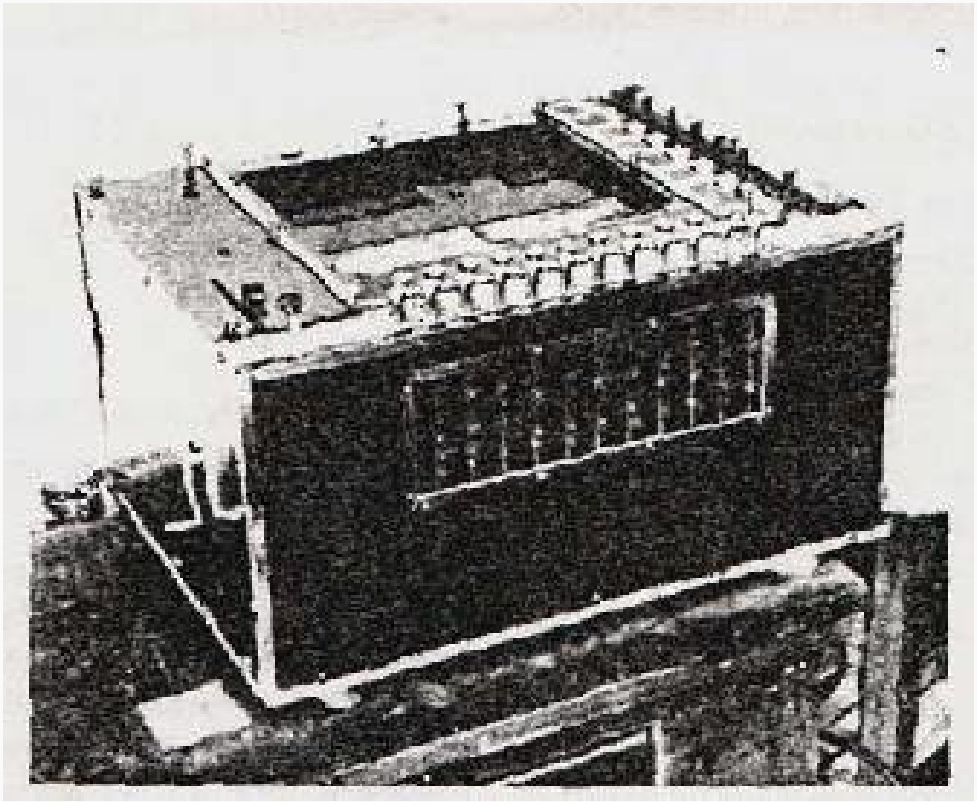
a word alignment



‘The Mathematics of Machine Translation: Parameter Estimation‘, Brown et al. (1993). ‘

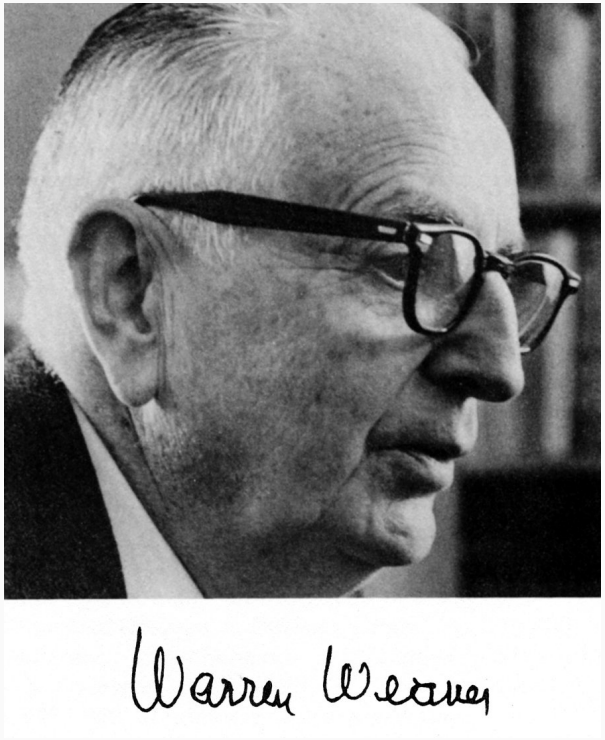
SOME OF THE STORY SO FAR

pre-history



1930s Peter Troyanskii and Georges Artsrouni patented mechanical translation devices.

ww2 cryptography



1940s Shannaon, Weaver, Turing: Information theory, Bayesian inference

great expectations

1954 Georgetown - IBM experiment Russian to English

Mi pyeryedayem mislyi posryedstvom ryechyi.
We transmit thoughts by means of speech.

- Translated 60 sentences.
- Claimed that MT would be solved within three or five years.
- Difference between limited and open domain.

disappointment

1966 ALPAC report

- Concluded that MT was too expensive and ineffective
- Recommended that research focus on tools to help human translators

statistical renaissance

1993 Brown et al., ‘The mathematics of statistical machine translation’

The Fundamental Equation of Statistical Machine Translation

$$\begin{aligned}\hat{\mathbf{e}} &= \operatorname{argmax} \mathbf{Pr}(\mathbf{e}|\mathbf{f}) \\ &= \operatorname*{argmax}_{\mathbf{e}} \mathbf{Pr}(\mathbf{e})\mathbf{Pr}(\mathbf{f}|\mathbf{e})\end{aligned}$$

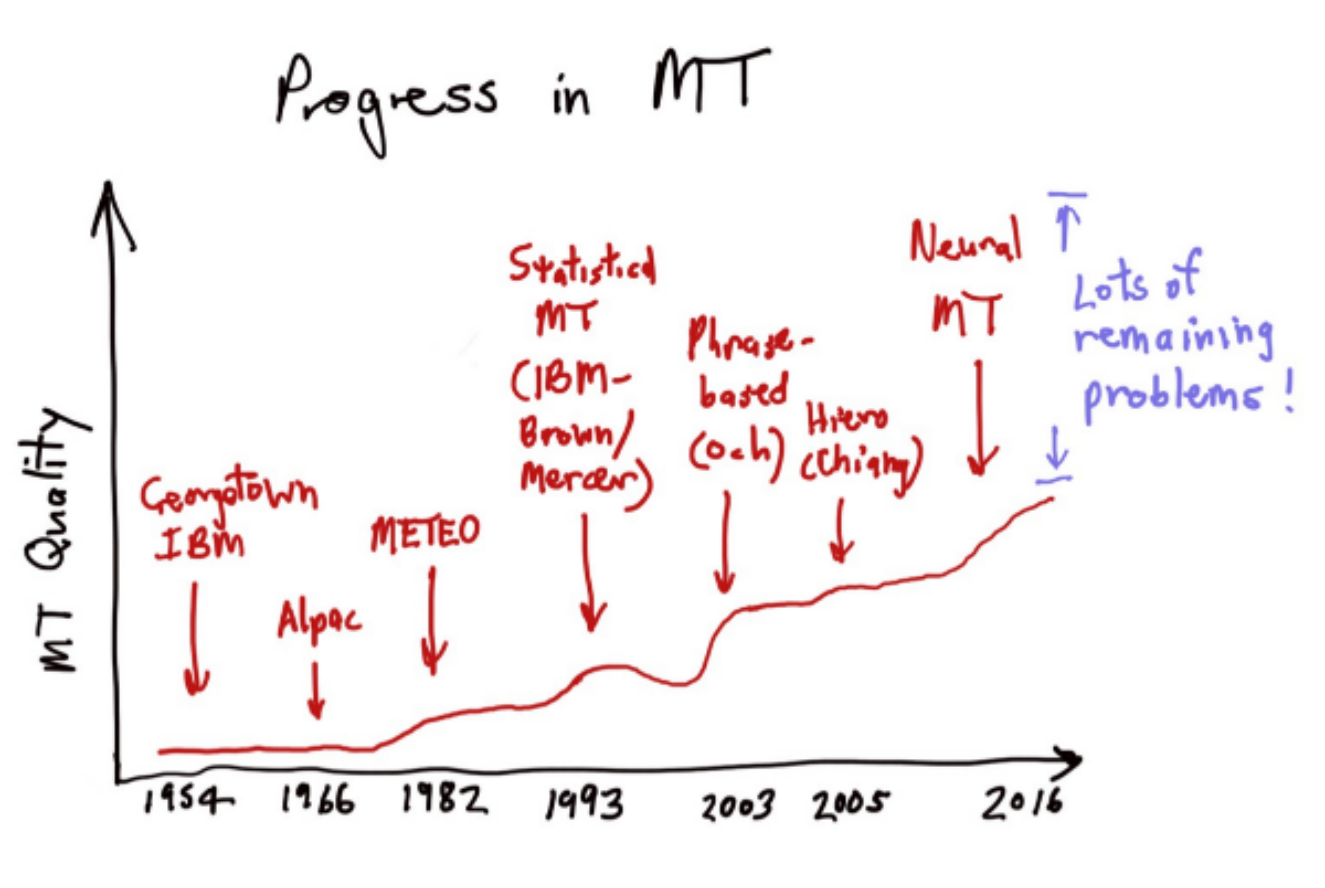
data

2000s Huge amounts of naturally occuring parallel data

gpus

GPUs with LSTMs and other robust recurrent neural networks

progress so far



(From Chris Manning's slides)

A PLAN

data driven machine translation

- Specify a simple statistical model of translation

data driven machine translation

- Specify a simple statistical model of translation
- Learn the parameters of the model from data

data driven machine translation

- Specify a simple statistical model of translation
- Learn the parameters of the model from data
- Use linguistic analysis to inform and constrain the model

parallel corpora

- Translated documents from governments, newspapers, etc.

parallel corpora

- Translated documents from governments, newspapers, etc.
- What's wrong with the data?

parallel corpora

- Translated documents from governments, newspapers, etc.
- What's wrong with the data?
  - It's often noisy

parallel corpora

- Translated documents from governments, newspapers, etc.
- What's wrong with the data?
  - It's often noisy
  - It's in the wrong domains (mostly)

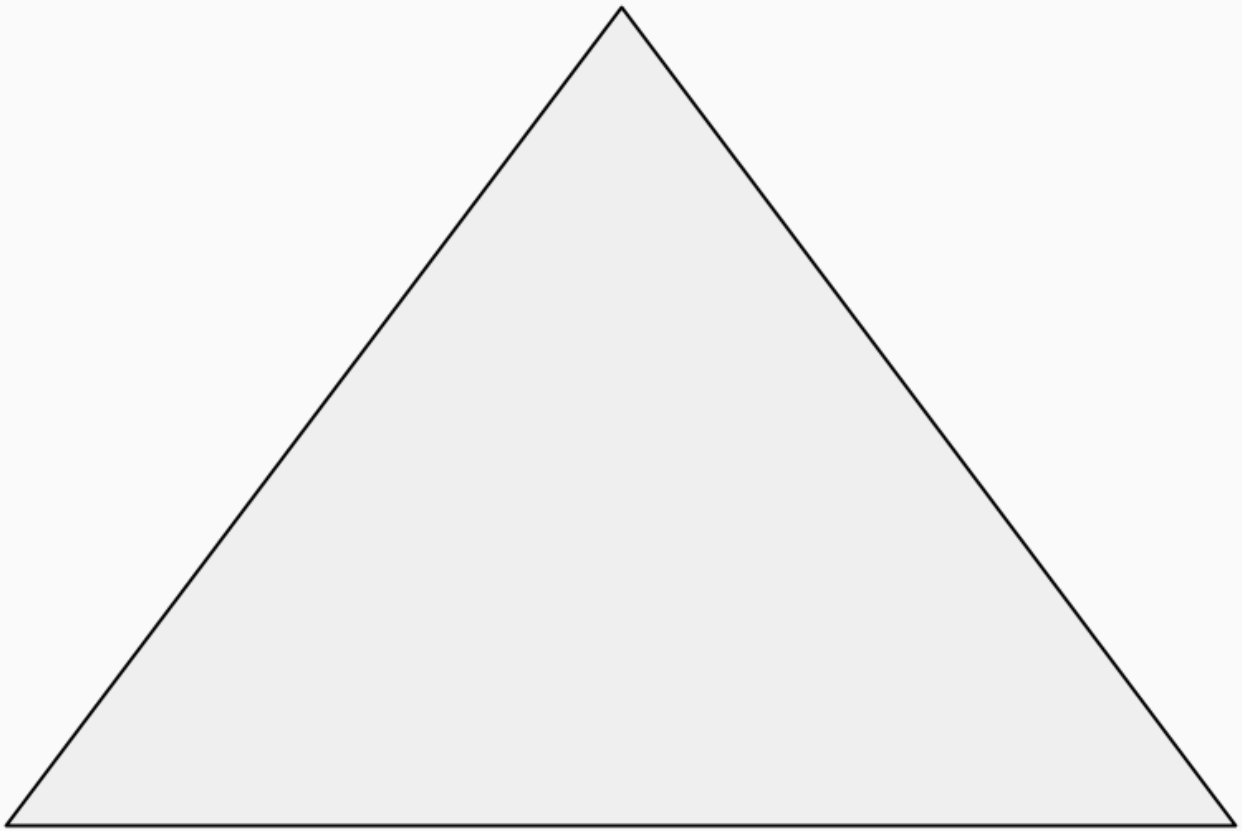
parallel corpora

- Translated documents from governments, newspapers, etc.
- What's wrong with the data?
  - It's often noisy
  - It's in the wrong domains (mostly)
  - It's only partially observed

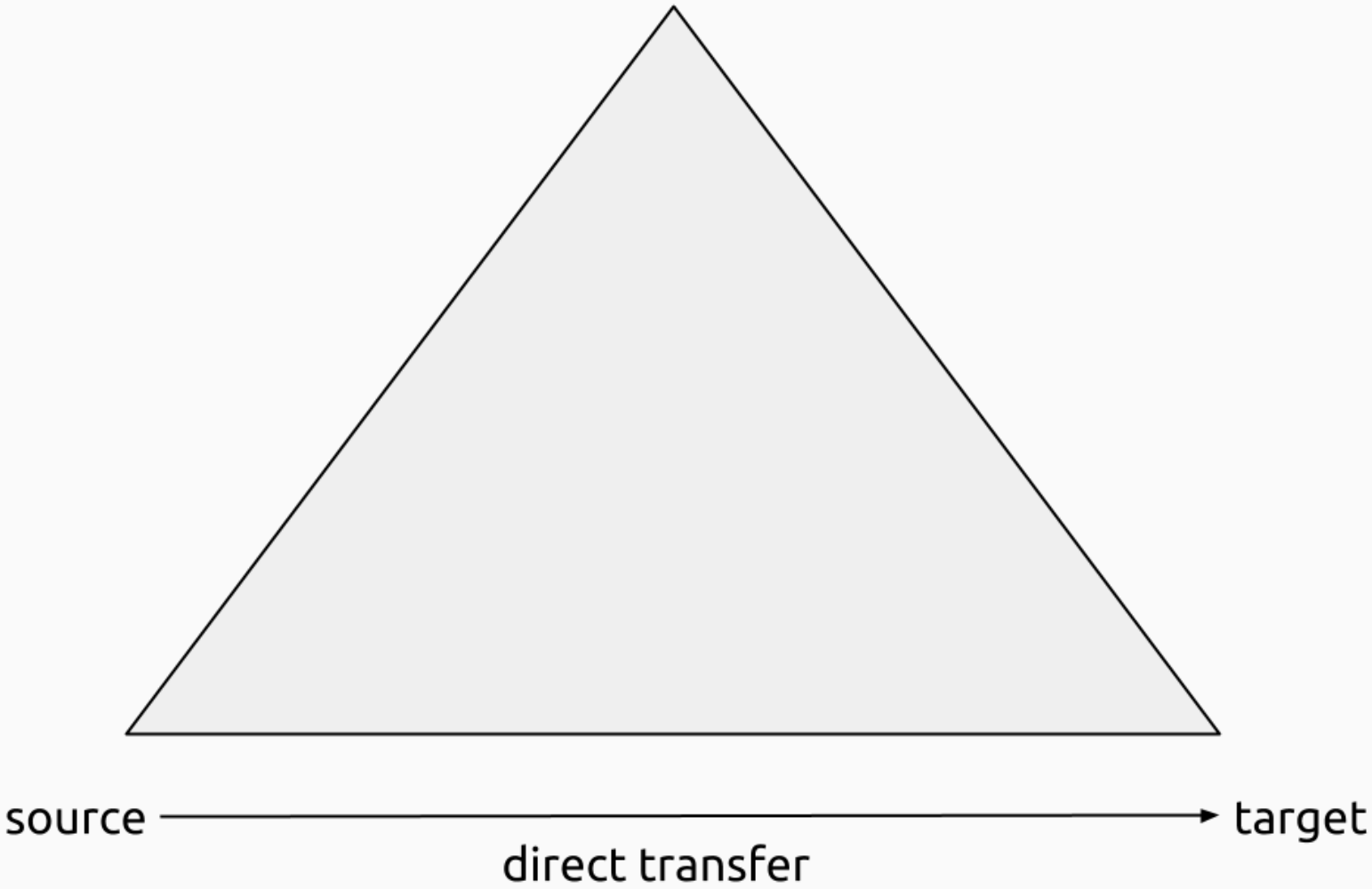
parallel corpora

- Translated documents from governments, newspapers, etc.
- What's wrong with the data?
  - It's often noisy
  - It's in the wrong domains (mostly)
  - It's only partially observed
  - There's not enough of it!

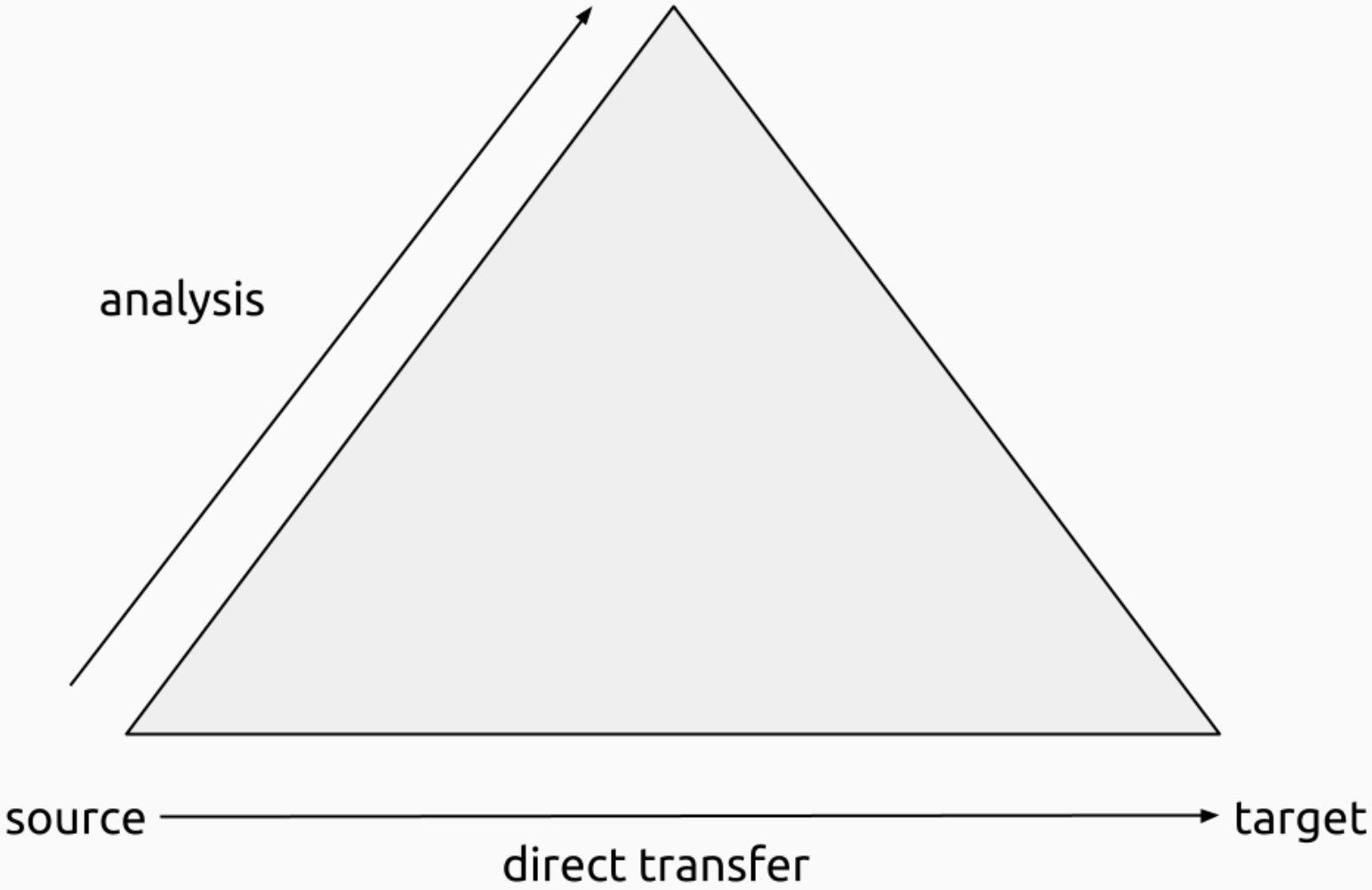
vauquois triangle



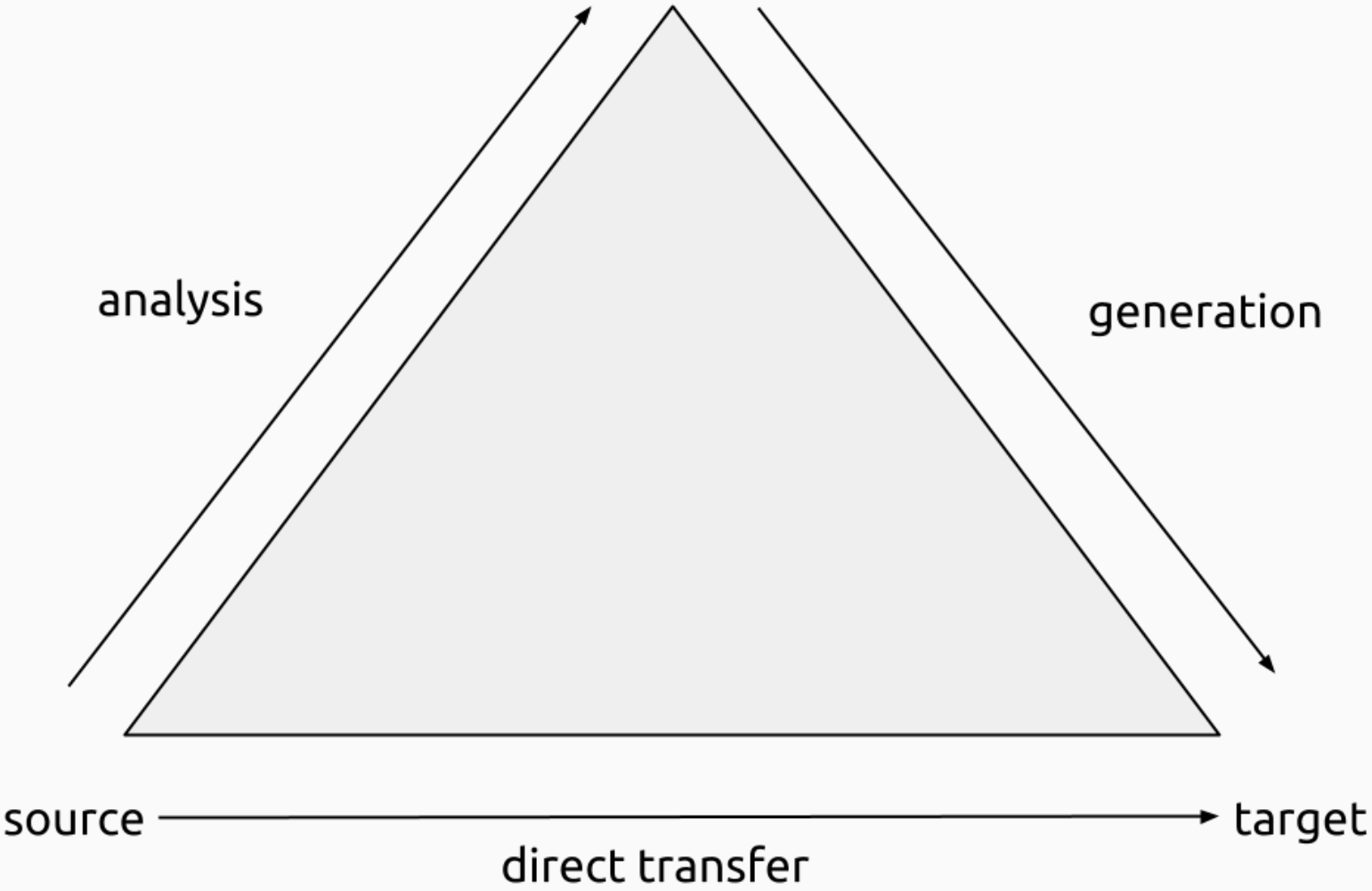
vauquois triangle



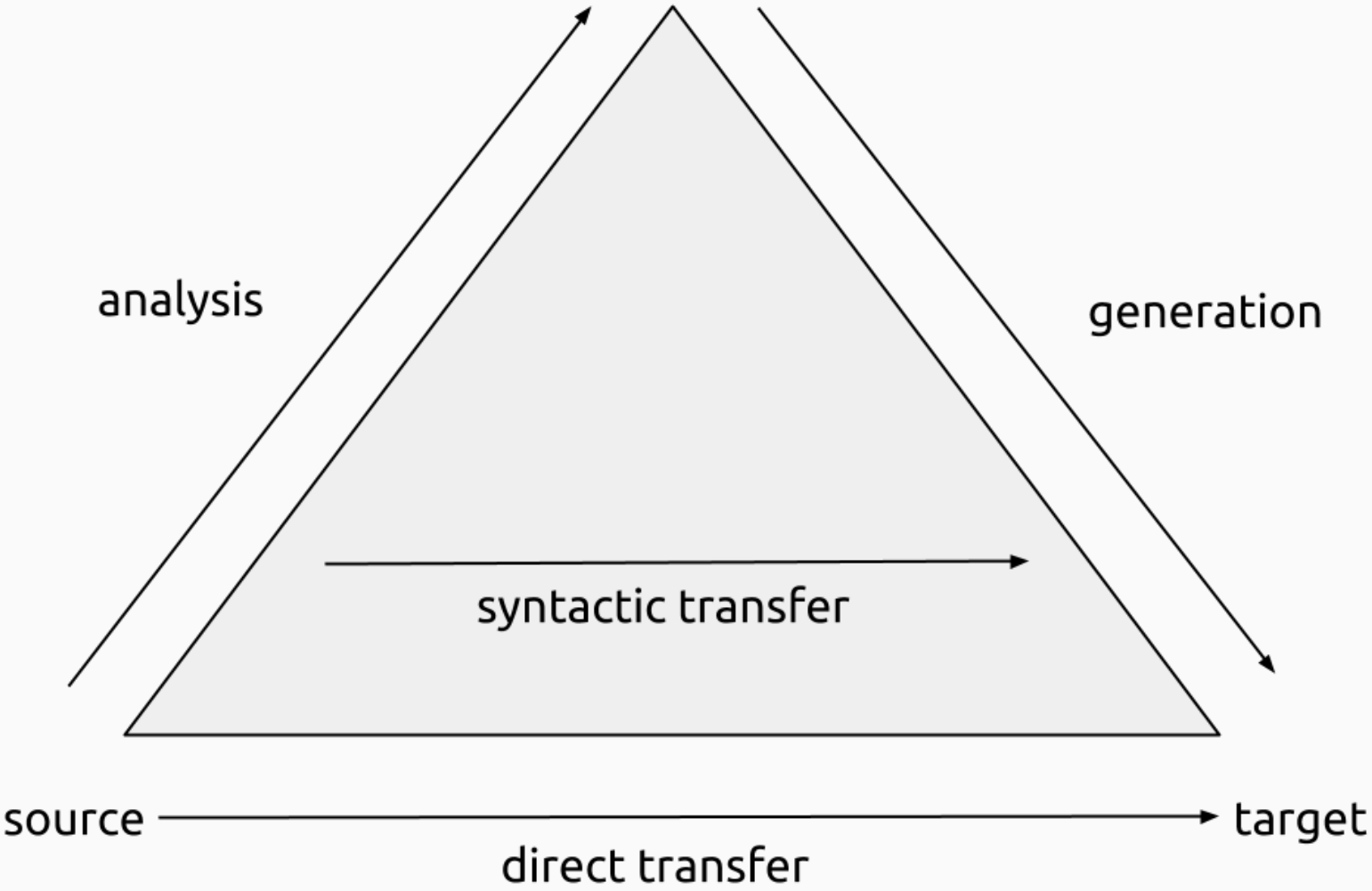
vauquois triangle



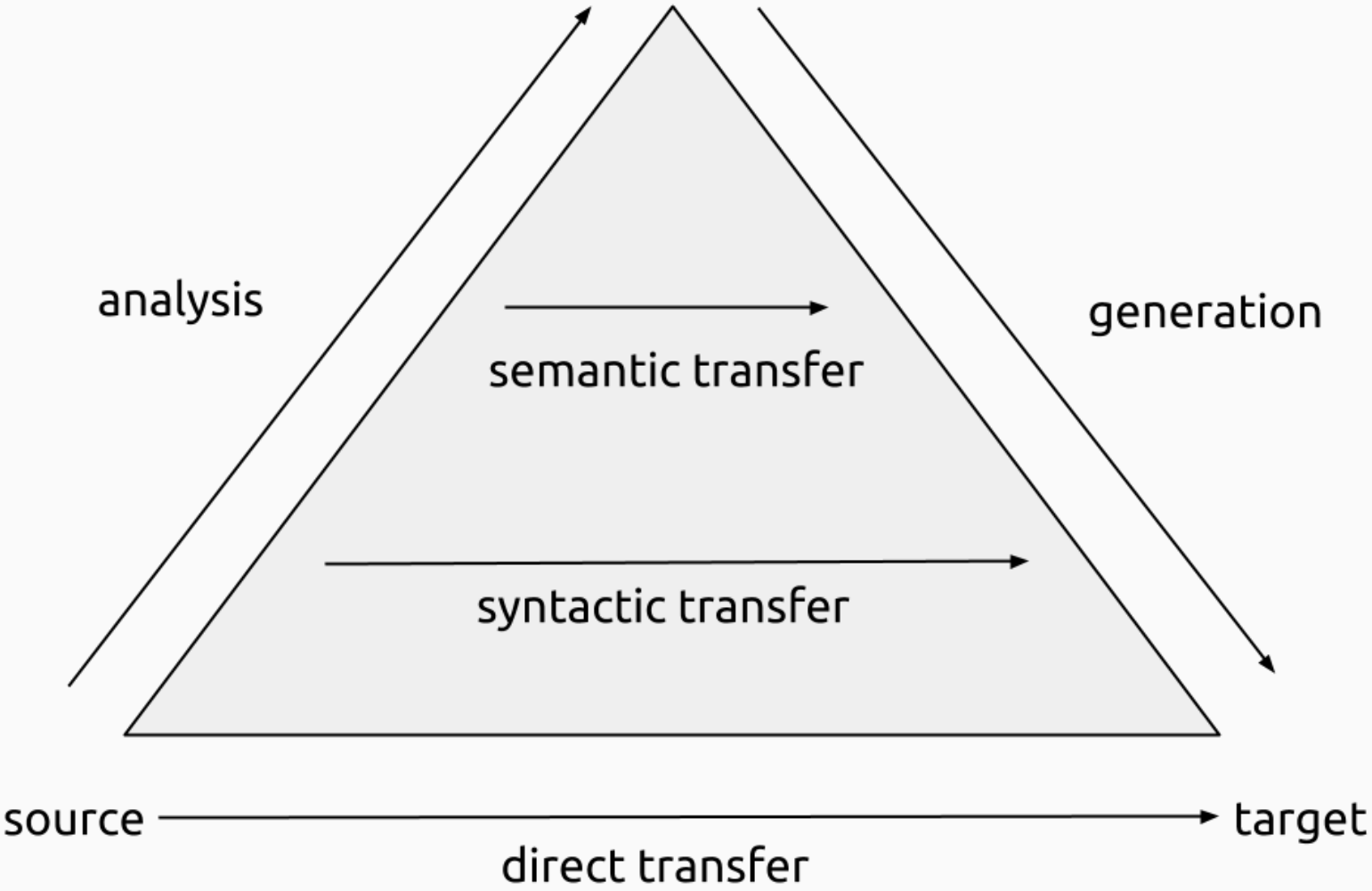
vauquois triangle



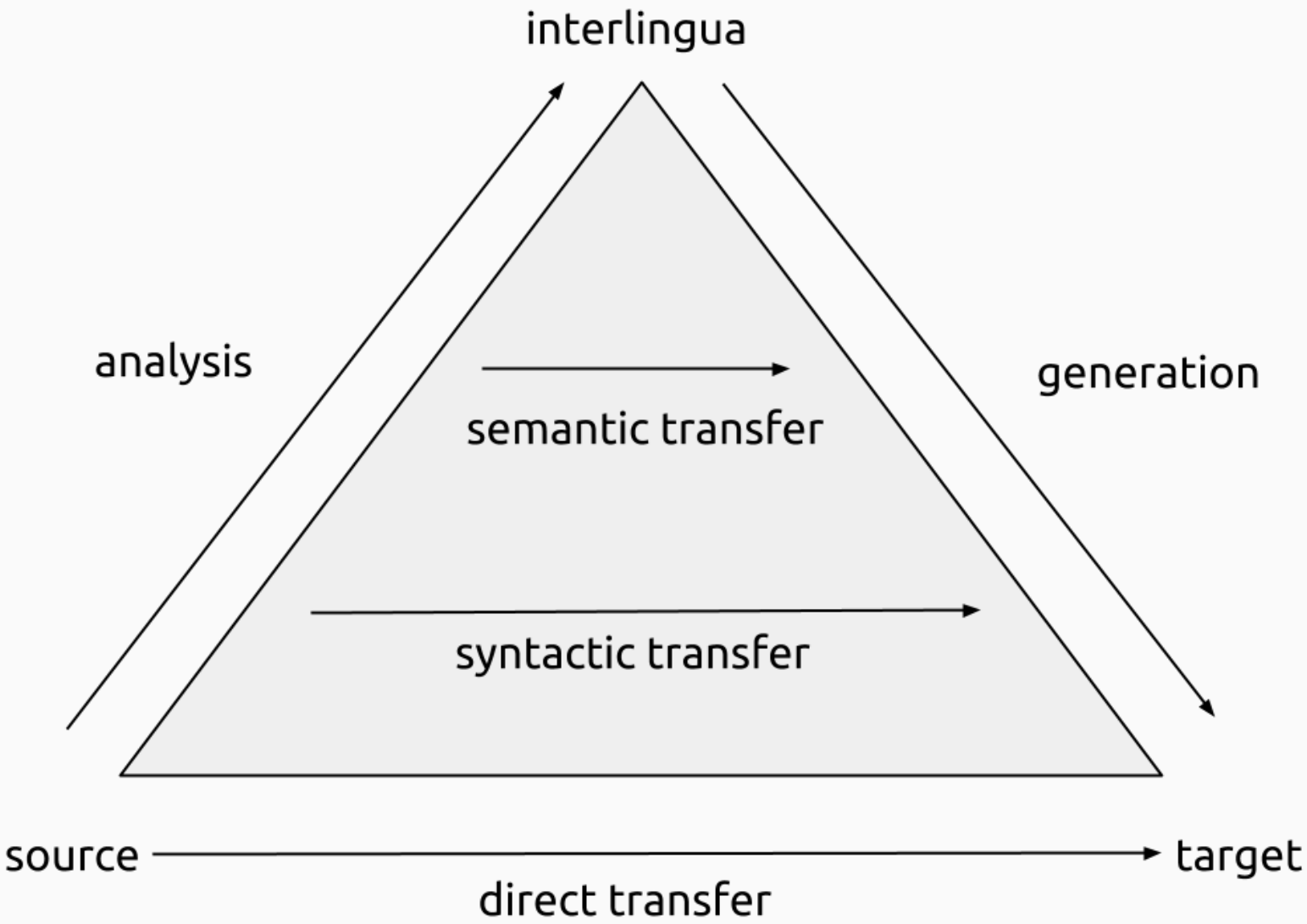
vauquois triangle



vauquois triangle



vauquois triangle



EVALUATION

hands-on evaluation

Compare Bing, Google and Yandex Translate

- Work in pairs
- Compare sentences from random wikipedia articles (en<->ru)
- Add the source sentence, translations and judgements to this spreadsheet: https://goo.gl/TcG5MZ