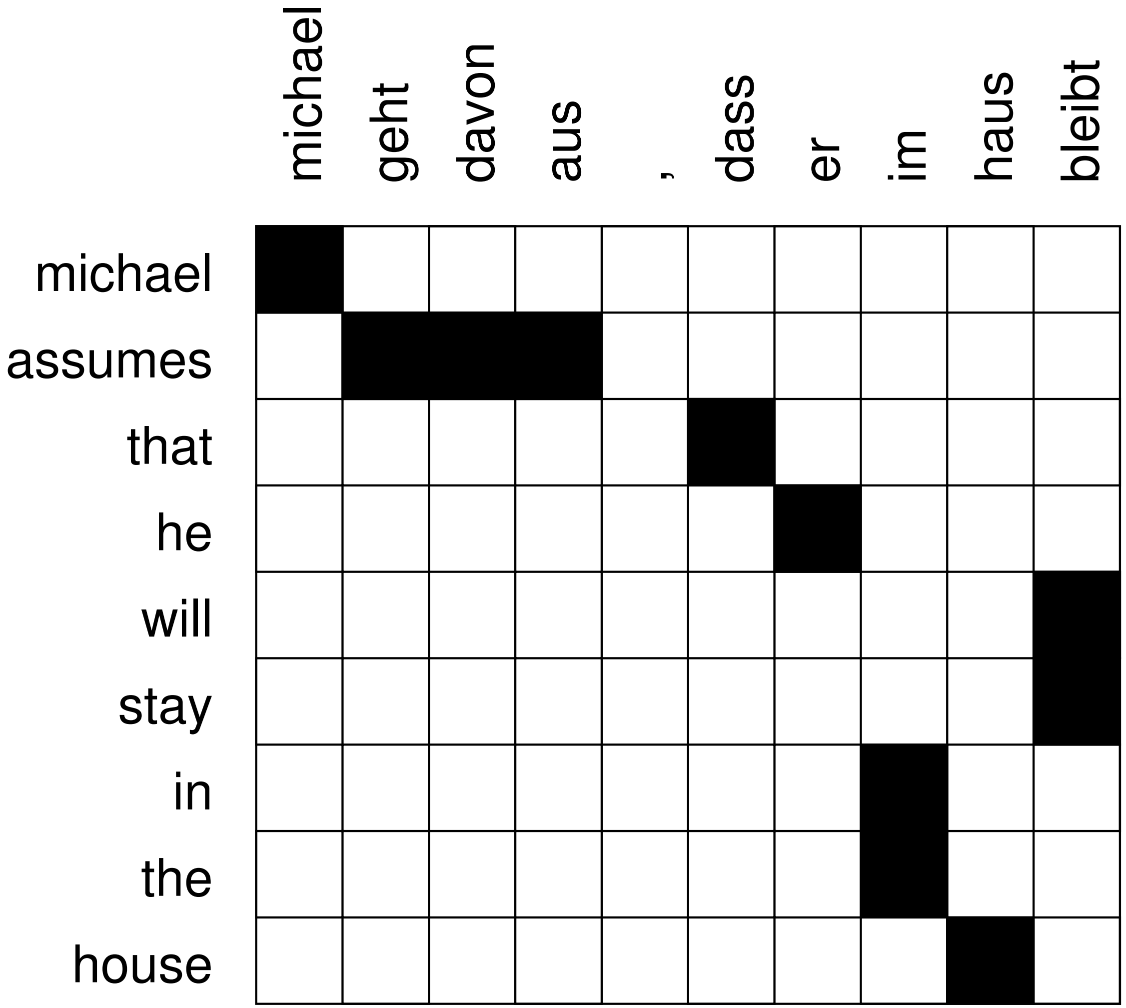
ENLP Lecture 21a
More on Machine Translation

Nathan Schneider

(with slides by Philipp Koehn, Chris Dyer)

29 November 2016

Word Alignment

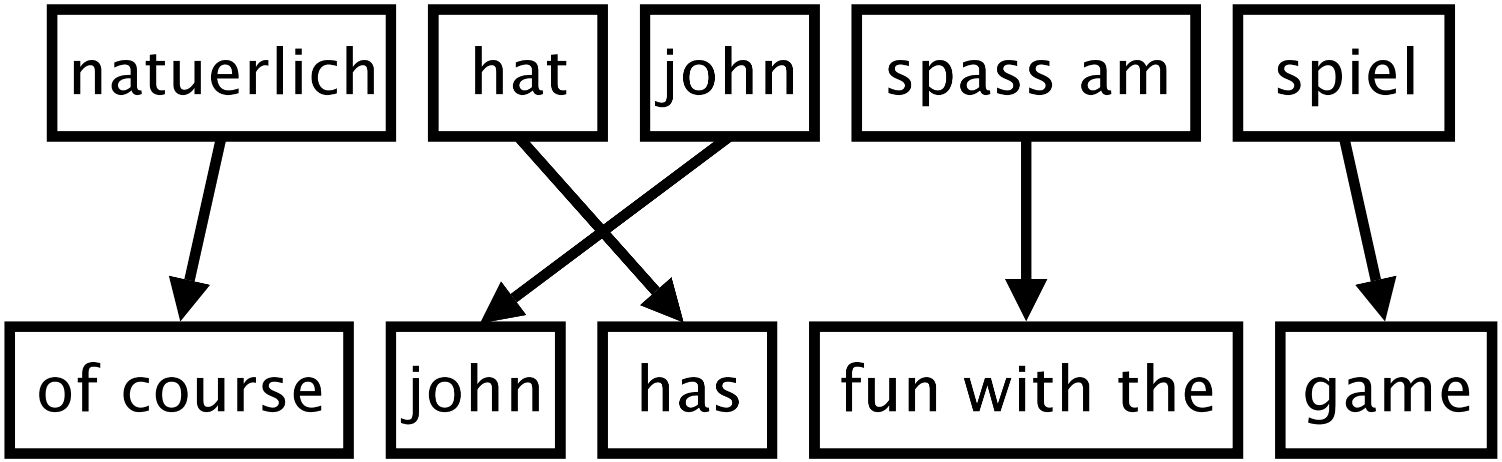


EM

- Procedure for optimizing generative models without supervision
  - Randomly initialize parameters, then
  - E: predict hidden structure ***y*** (hard or soft)

    M: estimate new parameters $\hat{\mathrm{P}}(\boldsymbol{y} \mid \boldsymbol{x})$ by MLE
- Likelihood function is non-convex. Consider trying several random initializations to avoid getting stuck in local optima.

Phrase-Based Model



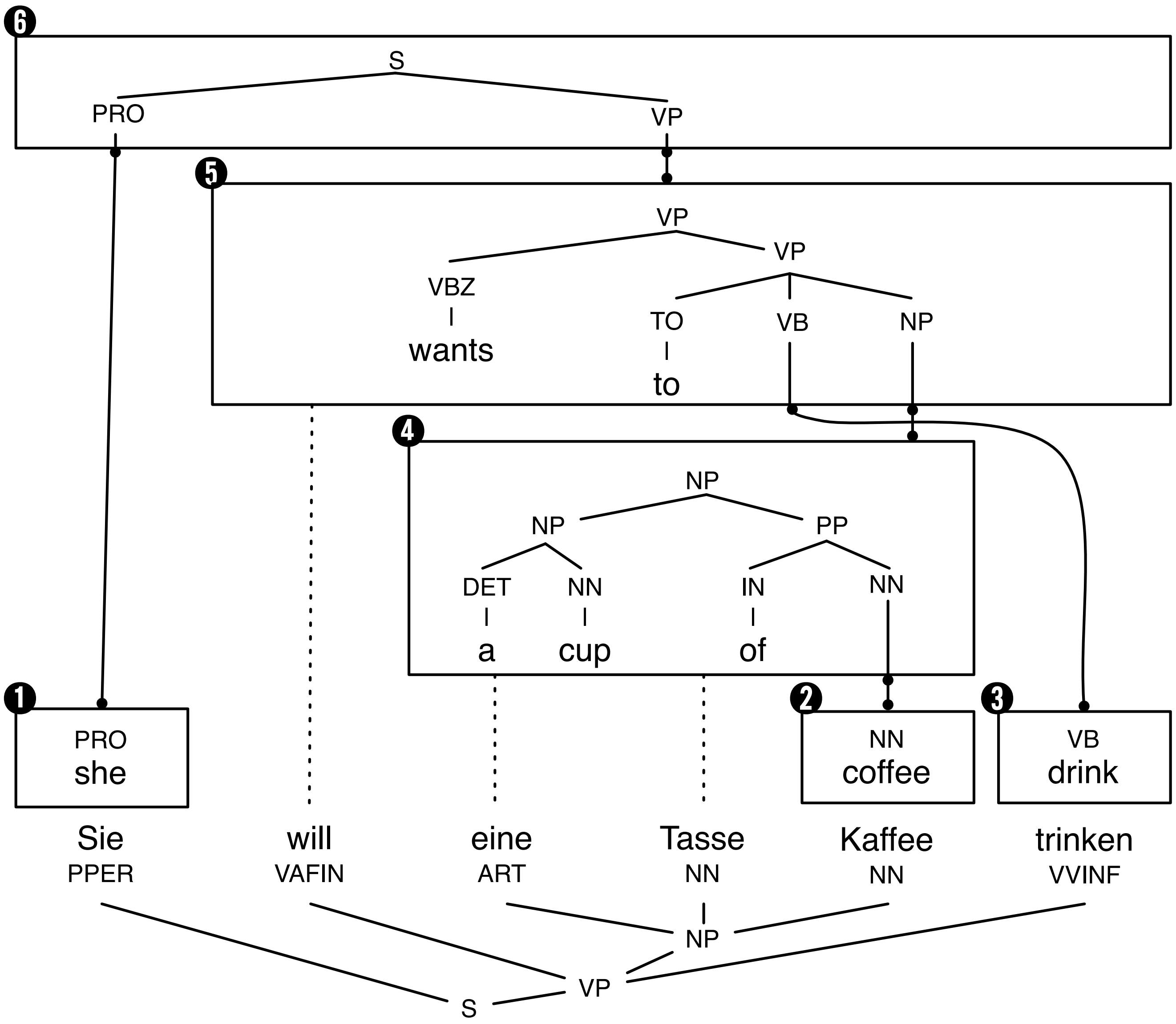
- Foreign input is segmented in phrases
- Each phrase is translated into English
- Phrases are reordered

- Workhorse of today's statistical machine translation

You as the Computer

в	этом	смысле	подобные	действия	частично	дискредитируют	систему	американской	демократии
in	this	sense	such	actions	some	discredit	system	american	democracy
the	that	meaning	similar	action	partially		a system	u.s.	democracies
a	the	terms	these	the	part		systems	us	democratic
at	it	way	this	acts	in part		which	america	of democracy
	here	sense ,	like	steps	partly		network	america's	
this			these actions					american democracy	
in this sense								america’s democracy	
in that sense								us democracy	
in this respect									

Syntax-Based Translation



Semantic Translation

- Abstract meaning representation [Knight et al., ongoing]

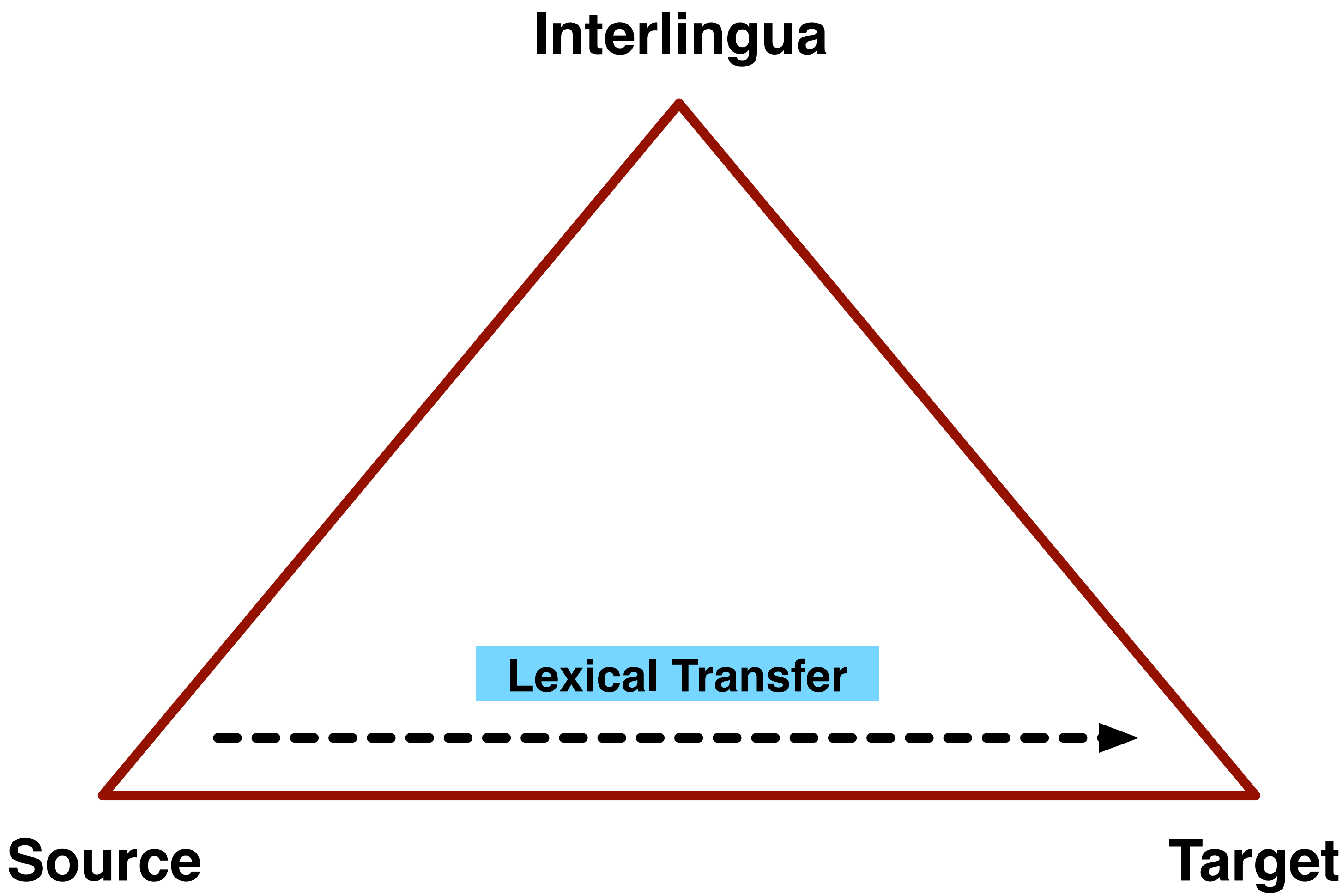
```
(w / want-01
   :agent (b / boy)
   :theme (l / love
       :agent (g / girl)
       :patient b))
```

- Generalizes over equivalent syntactic constructs (e.g., active and passive)

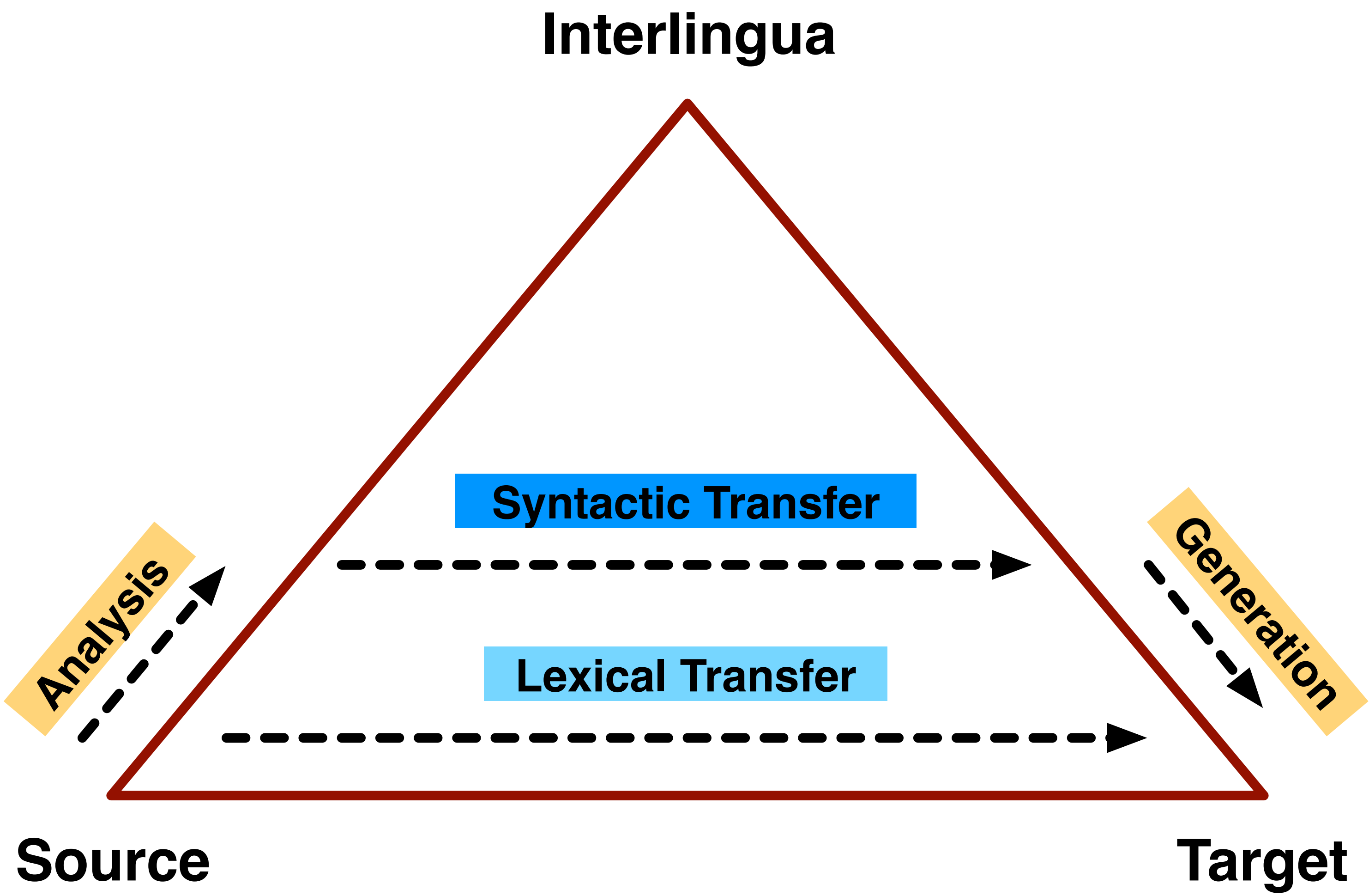
- Defines semantic relationships
  - semantic roles
  - co-reference
  - discourse relations

- In a very preliminary stage

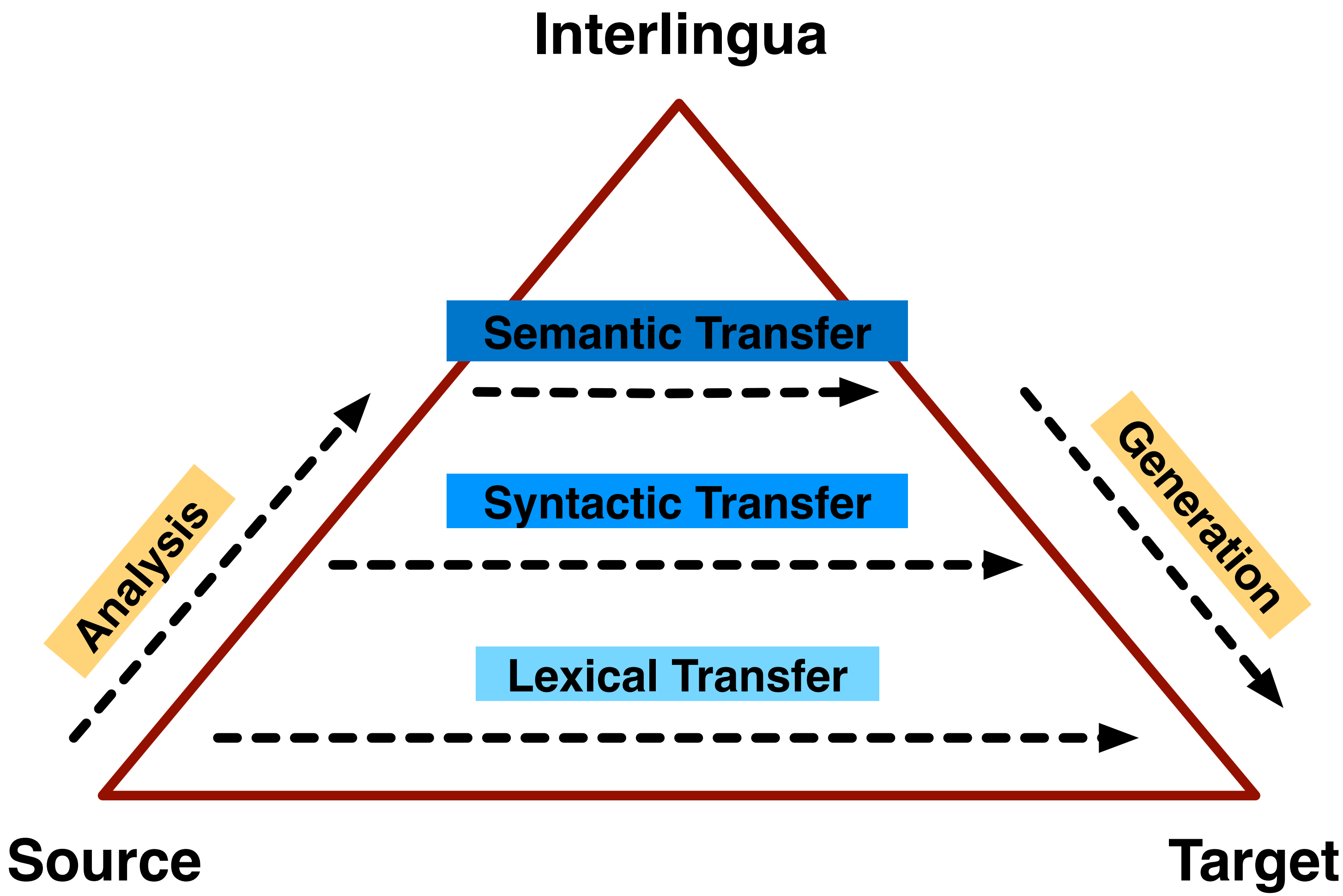
A Clear Plan



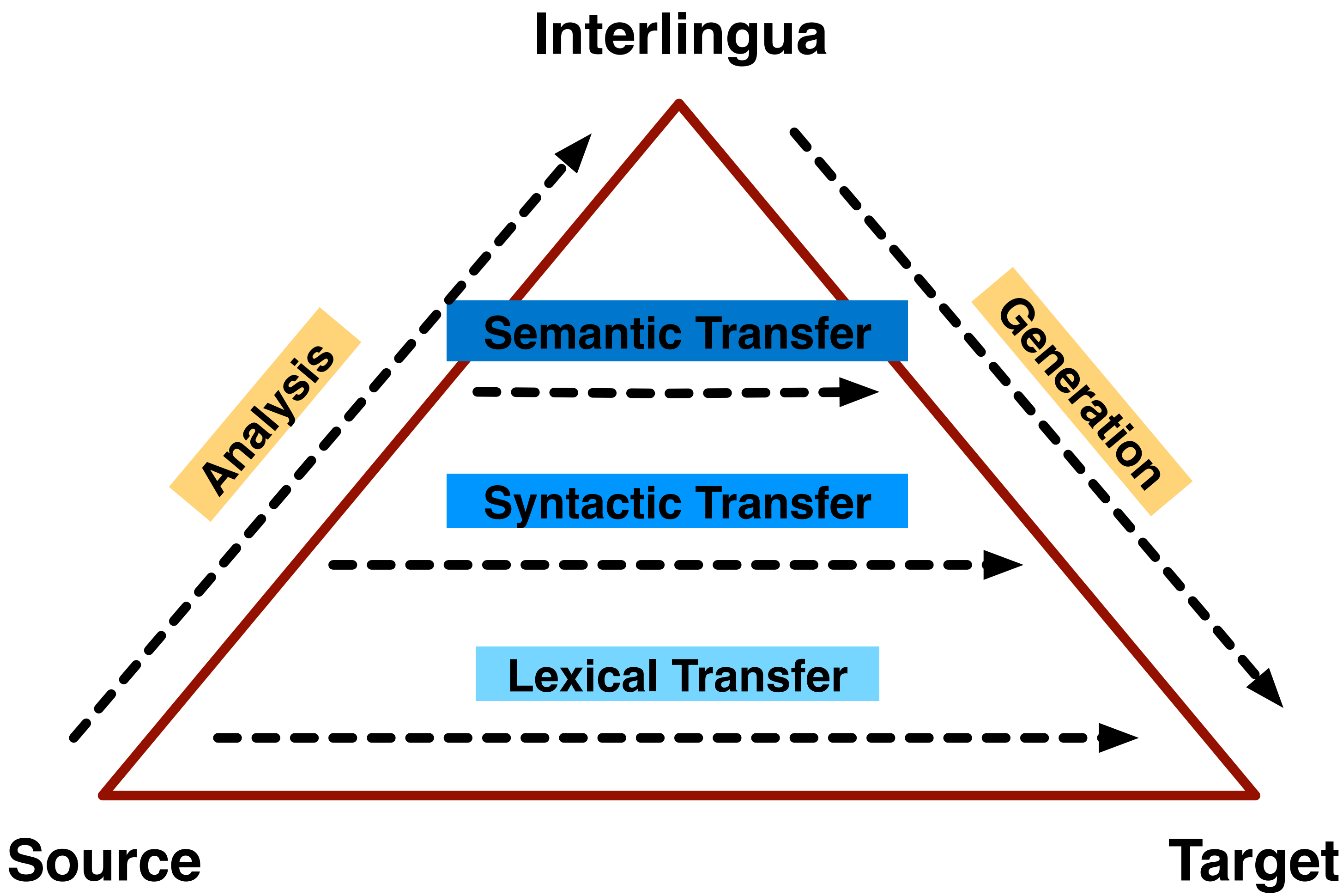
A Clear Plan


Interlingua
Syntactic Transfer
Lexical Transfer
Analysis
Generation
Source
Target

A Clear Plan


Interlingua
Semantic Transfer
Syntactic Transfer
Lexical Transfer
Analysis
Generation
Source
Target

A Clear Plan


Interlingua
Analysis
Generation
Semantic Transfer
Syntactic Transfer
Lexical Transfer
Source
Target

Neural MT

- Current research on neural network architectures, with state-of-the-art scores for some language pairs

Problem: No Single Right Answer

这个 机场 的 安全 工作 由 以色列 方面 负责 .

Israeli officials are responsible for airport security.
Israel is in charge of the security at this airport.
The security work for this airport is the responsibility of the Israel government.
Israeli side was in charge of the security of this airport.
Israel is responsible for the airport's security.
Israel is responsible for safety work at this airport.
Israel presides over the security of the airport.
Israel took charge of the airport security.
The safety of this airport is taken charge of by Israel.
This airport's security is the responsibility of the Israeli security officials.

Evaluation

Human Evaluation

- Manually **score** or **rank** candidate translations
  - e.g., for **fluency** (target language grammaticality/ naturalness) and **adequacy** (respecting the meaning of the source sentence)
- Manually **edit** the system output until it is an acceptable reference translation (HTER = Human Translation Edit Rate)
  - insertions, substitutions, deletions, shifts (moving a word or phrase)
  - then measure # edits / # words in reference (i.e., 1 – recall)

Automatic evaluation

- Why **automatic evaluation** metrics?
  - Manual evaluation is *too slow*
  - Evaluation on large test sets *reveals minor improvements*
  - **Automatic tuning** to improve machine translation performance

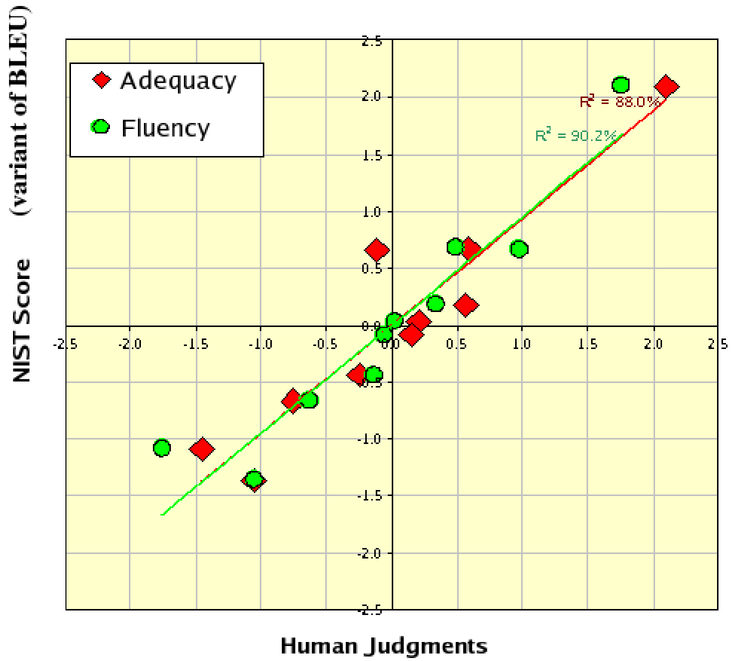
- History
  - Word Error Rate
  - **BLEU** since 2002

- BLEU in short: *Overlap with reference* translations

Automatic evaluation

- Reference Translation
  - the gunman was shot to death by the police .
- System Translations
  - **the gunman was police kill .**
  - **wounded police jaya of**
  - **the gunman was shot dead by the police .**
  - **the gunman arrested by police kill .**
  - **the gunmen were killed .**
  - **the gunman was shot to death by the police .**
  - **gunmen were killed by police ?SUB>0 ?SUB>0**
  - **al by the police .**
  - **the ringer is killed by the police .**
  - **police killed the gunman .**
- Matches
  - green = 4 gram match (good!)
  - red = word not matched (bad!)

Automatic evaluation



[from George Doddington, NIST]

- BLEU **correlates** with human judgement
  - **multiple reference translations** may be used

what is it good for?

what is it good *enough* for?

Quality

HTER	**assessment**
0%	
	publishable
10%	
	editable
20%	
30%	gistable
40%	triagable
50%	

(scale developed in preparation of DARPA GALE programme)

Applications

HTER	**assessment**	**application examples**
0%		Seamless bridging of language divide
	publishable	Automatic publication of official announcements
10%		
	editable	Increased productivity of human translators
20%		Access to official publications
		Multi-lingual communication (chat, social networks)
30%	gistable	Information gathering
		Trend spotting
40%	triagable	Identifying relevant documents
50%		

Current State of the Art

HTER	**assessment**	**language pairs and domains**
0%		
	publishable	French-English restricted domain
10%		French-English technical document localization
	editable	French-English news stories
20%		
		English-German news stories
30%	gistable	English-Czech open domain
40%	triagable	
50%		

(informal rough estimates by presenter)

Want to become an MT pro?

- MT course planned for **Spring 2018**; will focus on statistical approaches, building MT systems with Moses

MT: Summary

- Human-quality machine translation is an **AI-complete** problem.
  - All the challenges of NL: ambiguity, flexibility (difficult to evaluate!), vocabulary & grammar divergences between languages, context
  - State-of-the-art now good *enough* to be useful/commercially successful for some language pairs and purposes.
- Tension: simplistic models + huge data, or linguistically savvy models + less data? MT systems can be word-level, phrase-based, syntax-based, semantics-based/interlingua (**Vauquois triangle**)
- Statistical methods, enabled by large **parallel corpora** and **automatic evaluations** (such as **BLEU**), are essential for broad coverage
  - Automatic **word alignment** on parallel data via EM (IBM models)
  - Noisy channel model: n-gram **language model** for target language + **translation model** that uses probabilities from word alignments
  - Open-source toolkits like Moses make it relatively easy to build your own MT system from data