Science Foundation Ireland For what's next

Engaging Content
Engaging People

A Brief History of Machine Translation

Andy Way
ADAPT Centre @ Dublin City University

The ADAPT Centre is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

Machine Translation ...

... is like decrypting a secret code

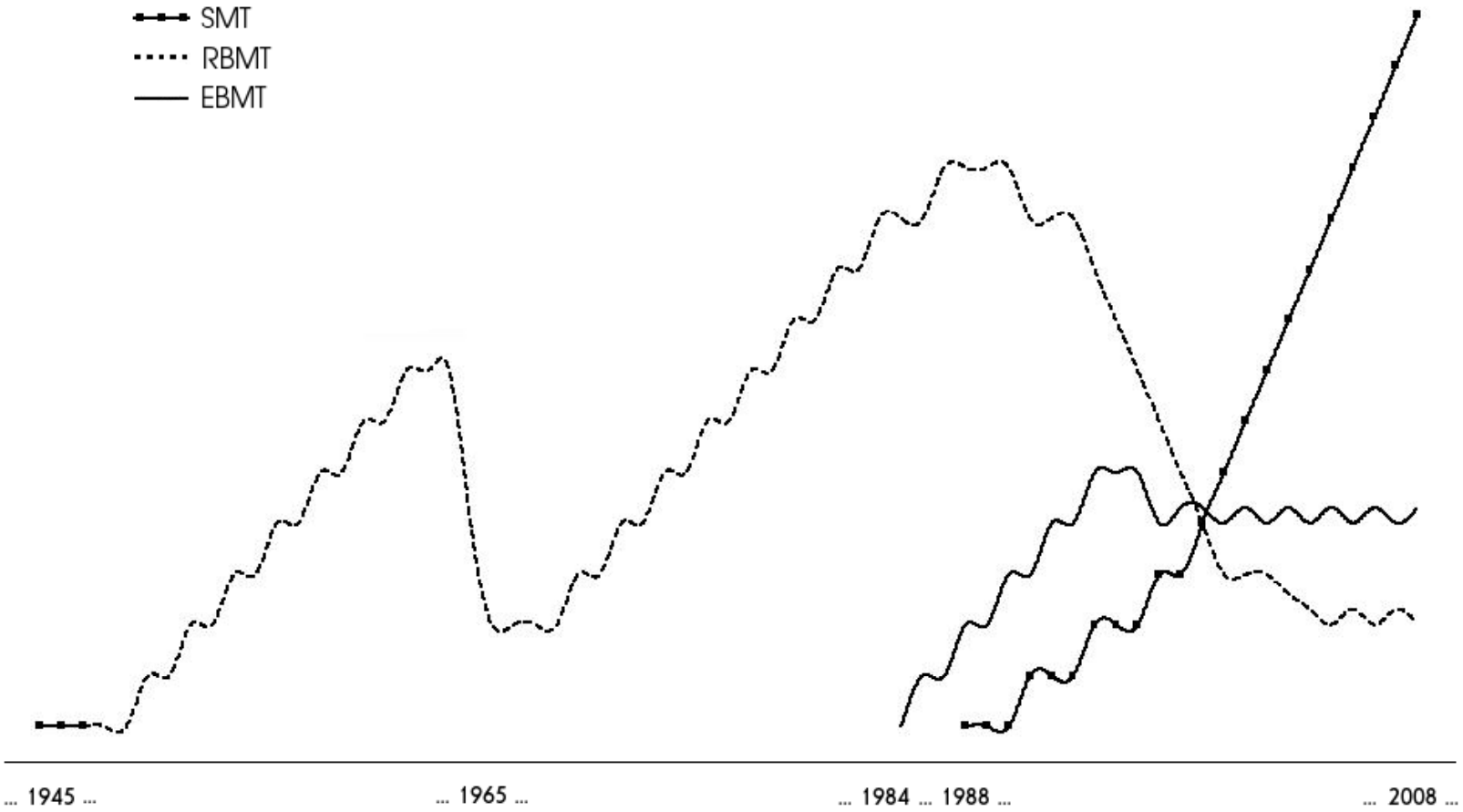
"Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography - methods which I believe succeed even when one does not know what language has been coded - one naturally wonders if the problem of translation could conceivably 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 (1894-1978)

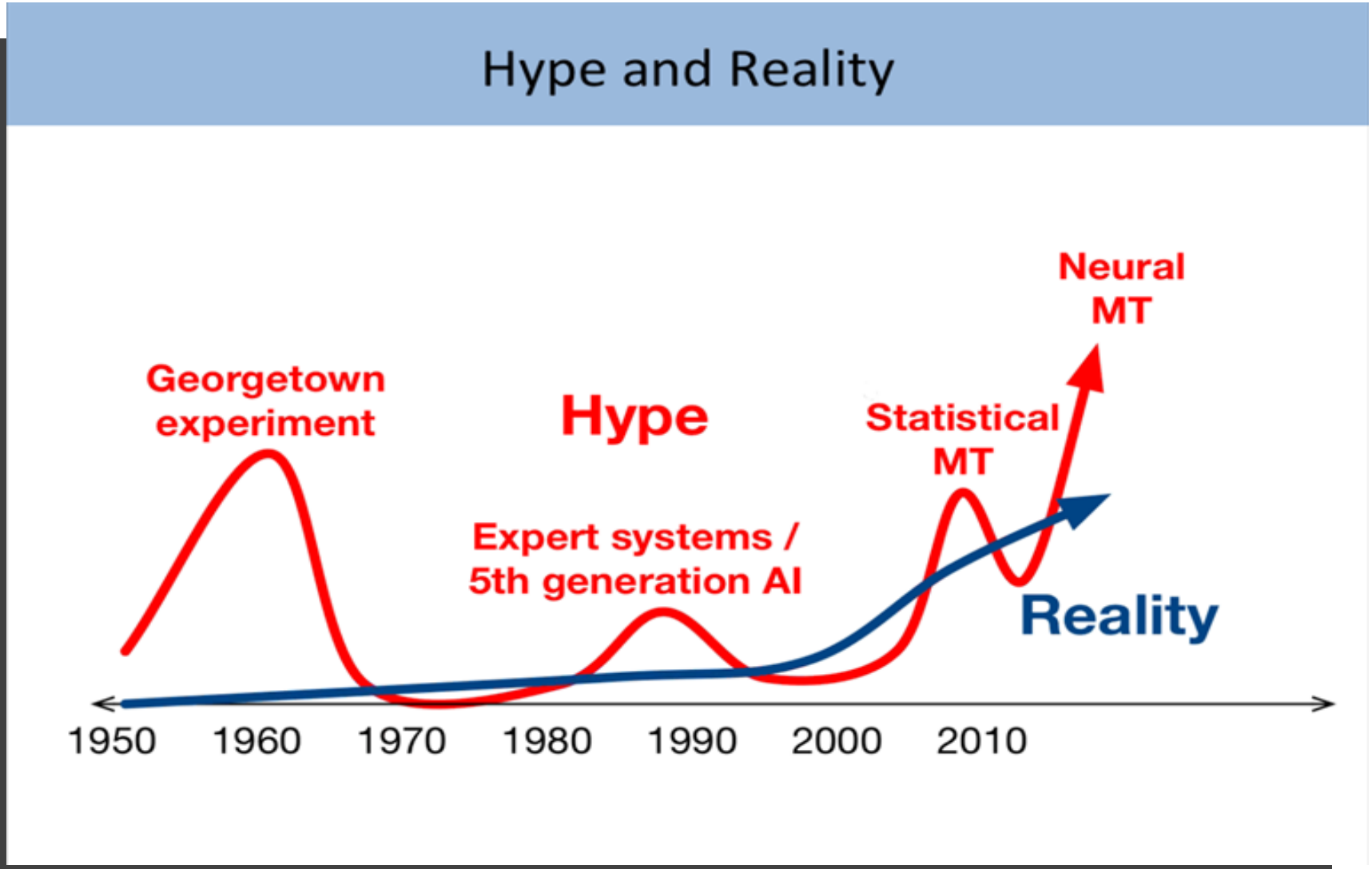
... computers learn how to translate by analyzing many human translations

Thanks to Marcello Federico

The Rise & Fall of Different MT Paradigms

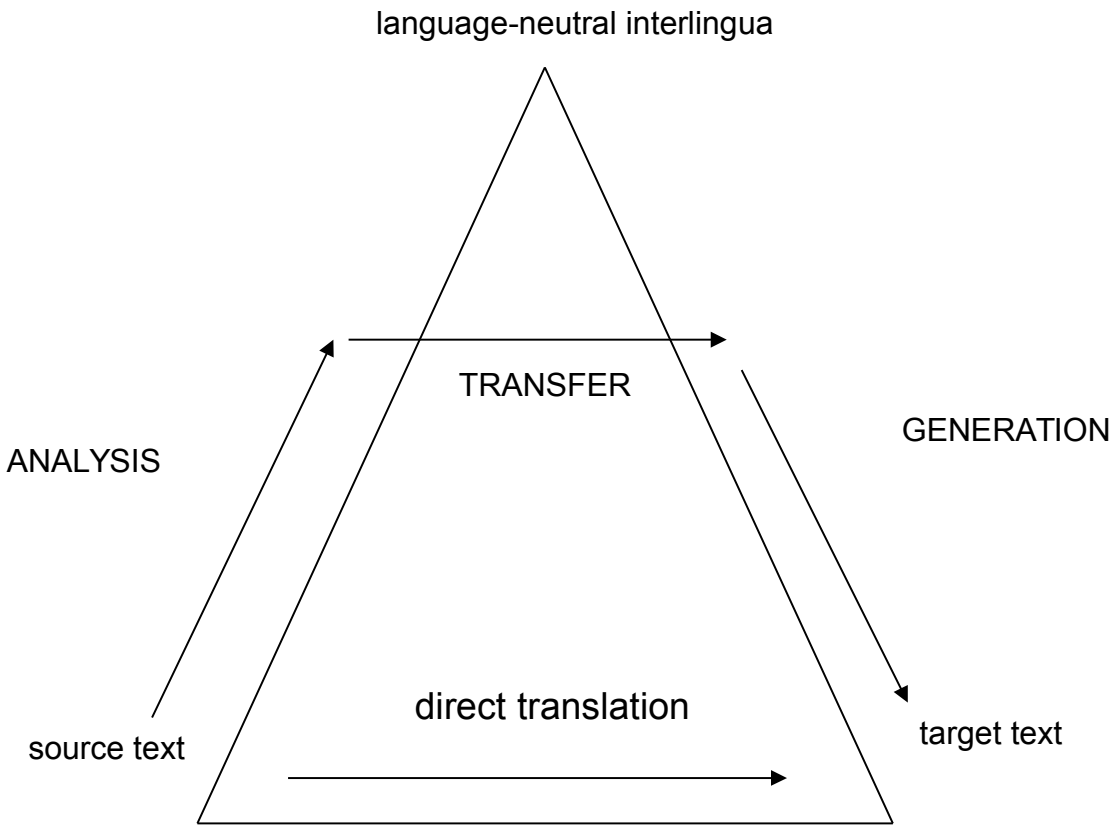


Another version ...



Thanks to Philipp Koehn: Omniscien Webinar 2017

Three main approaches to Rule-Based MT



The Vauquois Pyramid

System Design: Concerns

- **Multilingual vs. Bilingual**
  - Multilingual:
    - Extreme: Eurotra, i.e. 72 language pairs
    - Modest: EN ✉DE,FR,ES, i.e. 3 language pairs
    - Intermediate: EN,FR,DE,ES,JP, but not all combinations
  - Bilingual:
    - Unidirectional vs. Bidirectional
      - EN✉FR or FR🔔EN
    - Reversible vs. Non-reversible
      - EN🔔FR, same EN,FR components for Analysis & Generation, and reversible transfer module
      - EN✉FR & FR✉EN, but different EN, FR components for Analysis & Generation, and different transfer modules. NB, lack of modularity …
- **Direct vs. Transfer vs. Interlingua**
- **Batch vs. Interactive**

Advantages/Disadvantages of Direct Systems

- Advantages
  - Engine's competence lies in its *comparative grammar*.
  - Highly *robust*. Does not break down or stop when it encounters unknown words, unknown grammatical constructs, or ill-formed Input
  - Designed for *unidirectional* translation between one pair of langs. Not conducive to genuine multilingual MT design.
- Disadvantages
  - ‘word-for-word' translation + local reordering = poor translation, using cheap bilingual dictionary & rudimentary knowledge of target language.
  - Linguistically, computationally naive. No analysis of internal structure of Input, especially w.r.t. the grammatical relationships between the main parts of sentences.

Advantages/Disadvantages of Interlingual Systems

- Advantages
  - Intermediate representation (IR) fully specified, i.e. no need to ‘look back' at Source in order to generate Target.
  - Easy to extend to other langs.
  - Built-in *back translation*: useful for testing.

- Disadvantages
  - How to define an Interlingua for closely related languages?
  - Truly universal Interlingua possible?

Advantages/Disadvantages of Transfer Systems

- Advantages
  - No language-independent representations: source IR specific to a particular lang., as is the target lang. IR.
  - So complexity of Analysis & Generation components much reduced …
  - Also, no necessary equivalence between source and target IRs *for the same language*!
- Disadvantages
  - Not so easy to extend to other languages: *n* analysis modules, n generation modules, *n* x *n*-1 transfer modules, i.e. not much less than n^2!!!

Direct or Indirect?

- Direct:
  - From manufacturer's viewpoint, better, as it's more robust …
- Indirect:
  - Falls over more easily.
  - Development phase can be trying.
  - Commercially, must be supplemented with techniques for dealing with unseen Input.
- What about Translation Quality?
  - Indirect systems clearly better *in principle*.
  - However, constructing MT engine requires considerable effort.
  - Direct systems can achieve good performance.
- Summary
  - Research: mostly Transfer-based, with rules automatically acquired from data
  - Industrially: highly-developed Direct systems survived until quite recently …

Other Material on RBMT

- Arnold, D. *et al*. (1994): *Machine Translation - An Introductory Guide*; NCC Blackwell, Oxford
- Hutchins, J. & H. Somers (1992): *An Introduction to MT*; Academic Press, London
- Trujillo, A. (1999): *Translation Engines*; Springer, London

- Newer books included:
  - Bowker, L. (2002): *Computer-Aided Translation Technology*, U. of Ottawa Press.
  - Somers, H. (2003): *Computers and Translation: A translator's guide*, John Benjamins.
  - Bond, F. (2005): *Translating the Untranslatable*, CSLI.
  - Quah, C. (2006): *Translation and Technology*, Palgrave MacMillan.

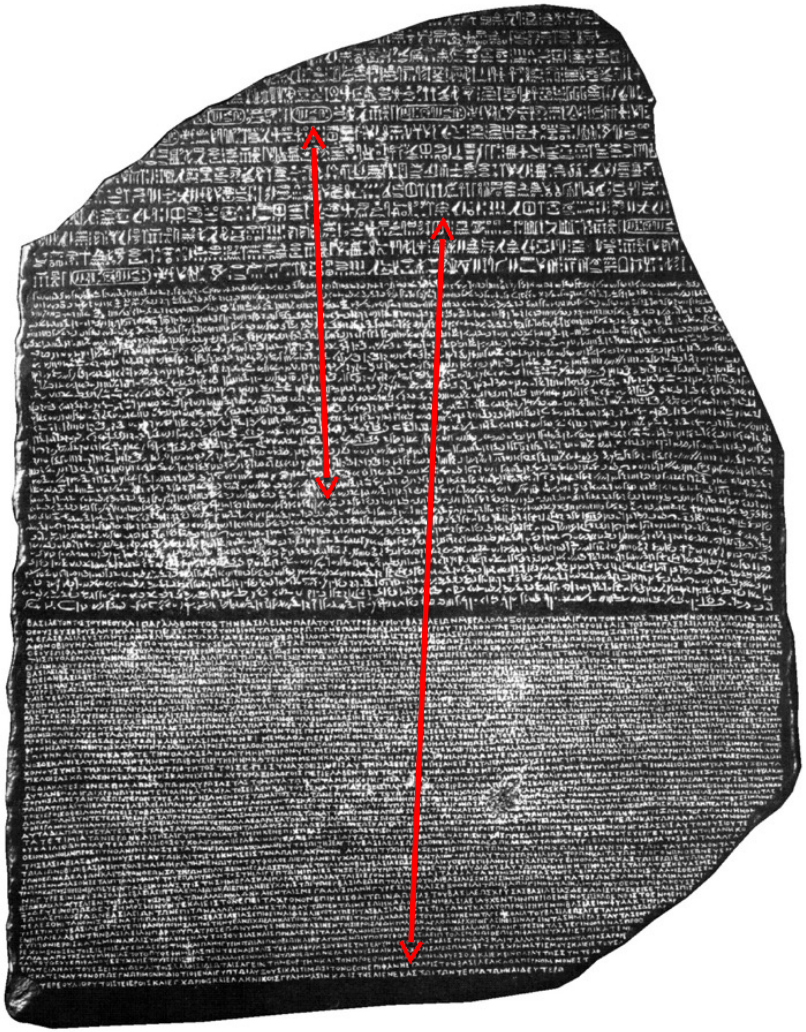
- Apertium: https://www.apertium.org/index.eng.html?dir=cat-spa#translation

Learning Translation

hieroglyphics

demotic

Greek



Egypt, 1799

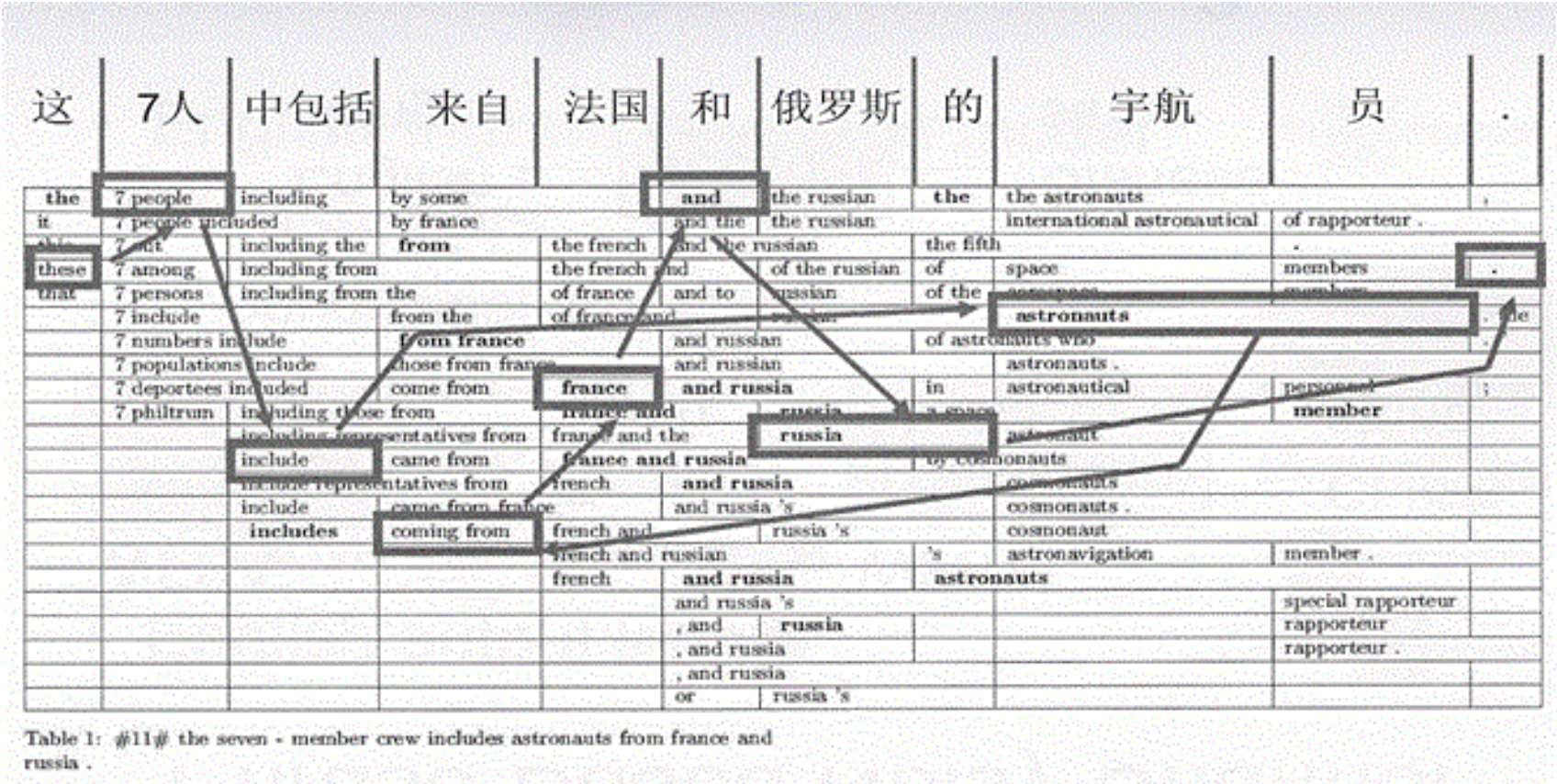
Why corpus-based approaches?

- the (relative) failure of rule-based approaches
- the increasing availability of machine-readable text
- the increase in capability of hardware (CPU, memory, disk space) with associated decrease in cost

SMT broken down into two problems

- **Search problem**:
  - how to efficiently generate/explore translation hypotheses
- **Modeling problem**:
  - how to score translation hypotheses

Phrase-Based Translation is Complex



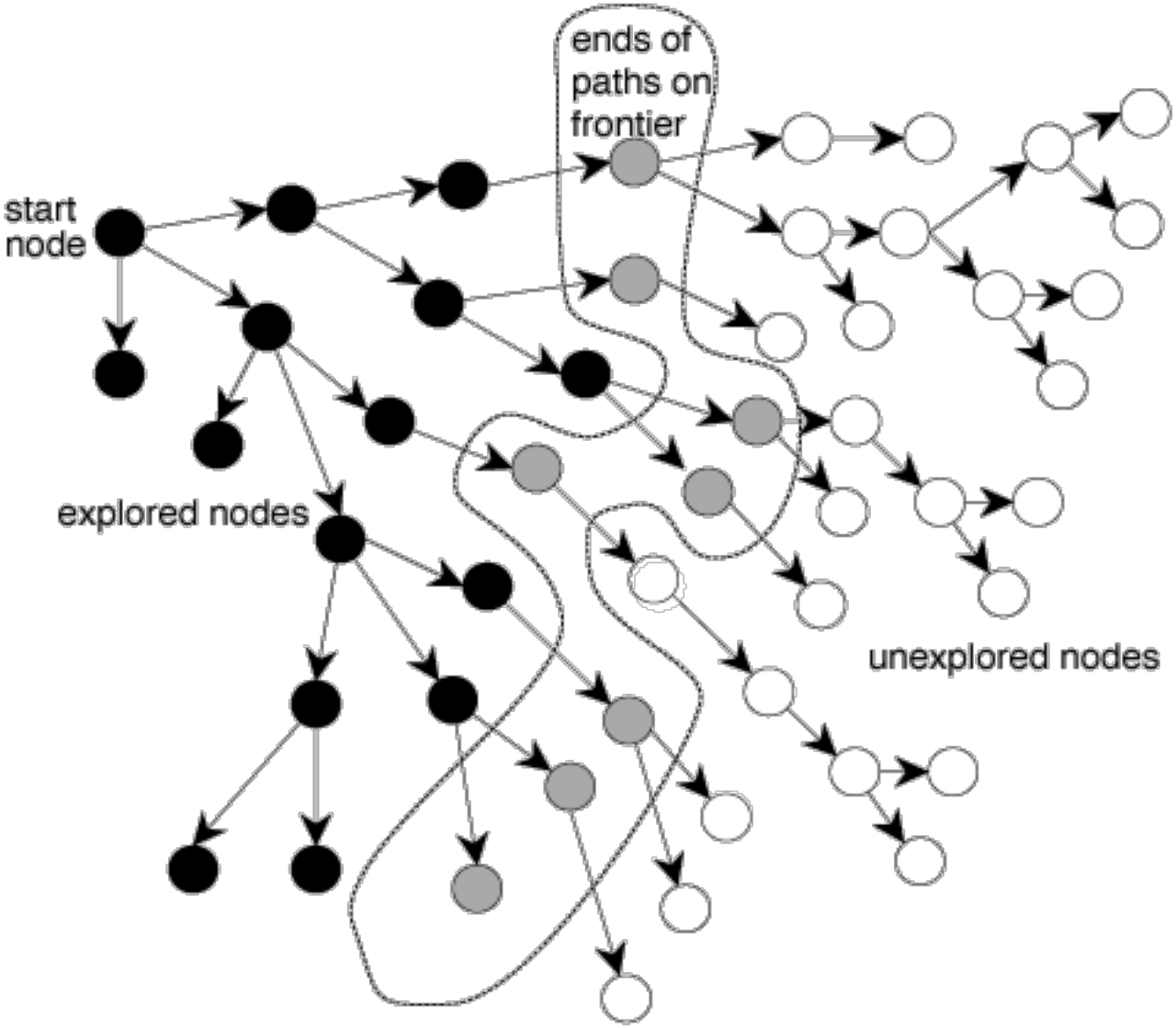
这	7人	中包括	来自	法国	和	俄罗斯	的	宇航	员	.
the	7 people	including	by some		and	the russian	the	the astronauts		
it	7 people included		by france		and the	the russian		international astronautical	of rapporteur	.
this	7 of	including the	from	the french	and the russian		the fifth			
these	7 among	including from		the french and		of the russian	of	space	members	.
that	7 persons	including from the		of france	and to	russian	of the			
	7 include		from the	of france and				astronauts		
	7 numbers include		from france		and russian		of astronauts who			
	7 populations include		those from france		and russian			astronauts .		
	7 deportees included		come from	france	and russia		in	astronautical	personnel	
	7 philtrum	including those from		france and		russia	a space		member	
		including representatives from		france and the		russia		astronaut		
		include	came from	france and russia			by cosmonauts			
		include representatives from		french	and russia			cosmonauts		
		include	came from france		and russia 's			cosmonauts .		
		includes	coming from	french and		russia 's		cosmonaut		
				french and russian			's	astronavigation	member .	
				french	and russia		astronauts			
					and russia 's				special rapporteur	
					, and	russia			rapporteur	
					, and russia				rapporteur .	
					, and russia					
					or	russia 's				

Table 1: #11# the seven - member crew includes astronauts from france and russia .

- Scoring:
  - Try to use phrase-pairs that have been frequently observed
  - Try to output a sentence with frequent English word sequences

Thanks to Kevin Knight

Search involves pruning hypotheses



Thanks to Marcello Federico

Not an Easy Time at the Beginning

P. Brown, J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer, P. Roosin: “A Statistical Approach to Language Translation”, COLING, 1988.

“**The validity of a statistical** (information theoretic) **approach to MT** has indeed been recognized, as the authors mention, by Weaver as early as 1949. And **was universally recognized as mistaken** by 1950 (cf. Hutchins, MT – Past, Present, Future, Ellis Horwood, 1986, p. 30ff and references therein). **The crude force of computers is not science**. The paper is simply beyond the scope of COLING.”

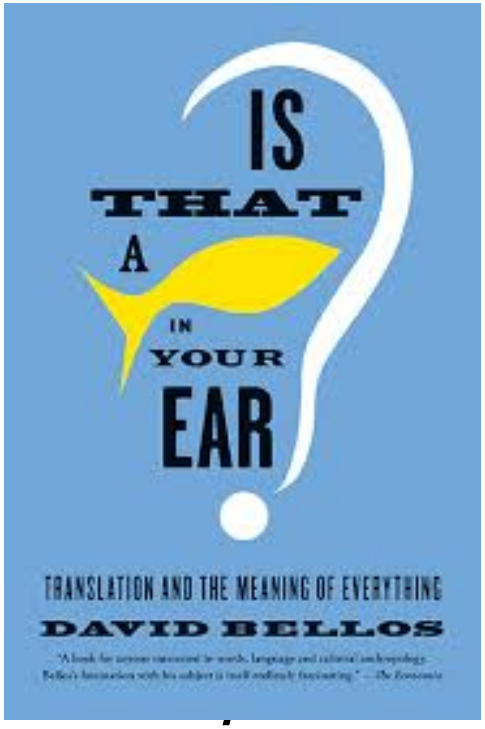
– Anonymous review

Other Material on SMT

- Koehn (2009): *Statistical Machine Translation*, CUP
- Goutte et al. (2009): *Learning Machine Translation*, MIT Press
- SMT tutorials from Knight, Koehn etc.
- MT Chapter in Jurafsky & Martin
- My Chapter on MT in *NLP Handbook* (Lappin et al. (eds) 2010)
- Carl & Way (2003): *Recent Advances in EBMT*, Kluwer
- Various conference proceedings (e.g. MT Summits, EAMT, AMTA …)

Other Material on SMT

- Koehn (2009): *Statistical Machine*
- Carl & Way (2003): *Recent Advances in EBMT*, Kluwer
- Various conference proceedings (e.g. MT Summits, EAMT, AMTA …)

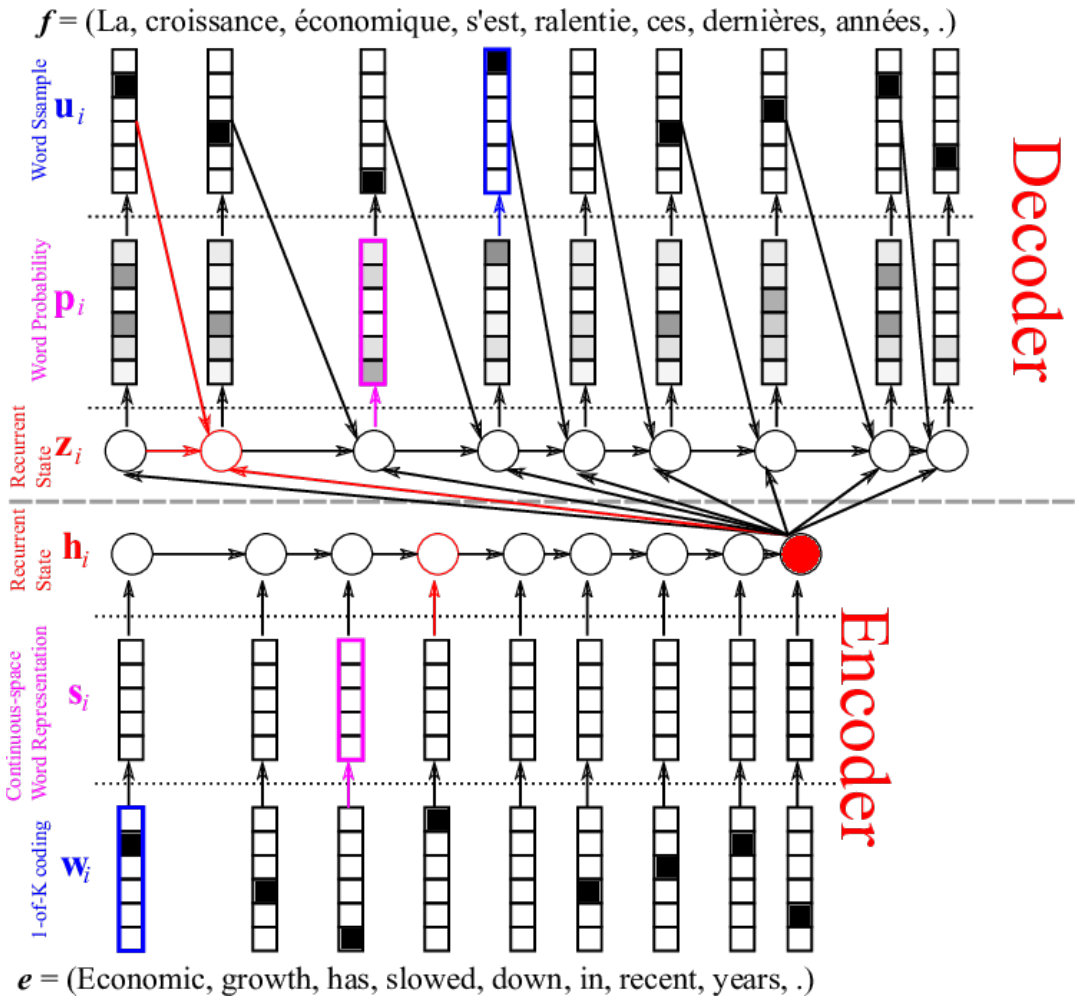


Or listen to a (respected!) translator …

SMT is **hard**!!

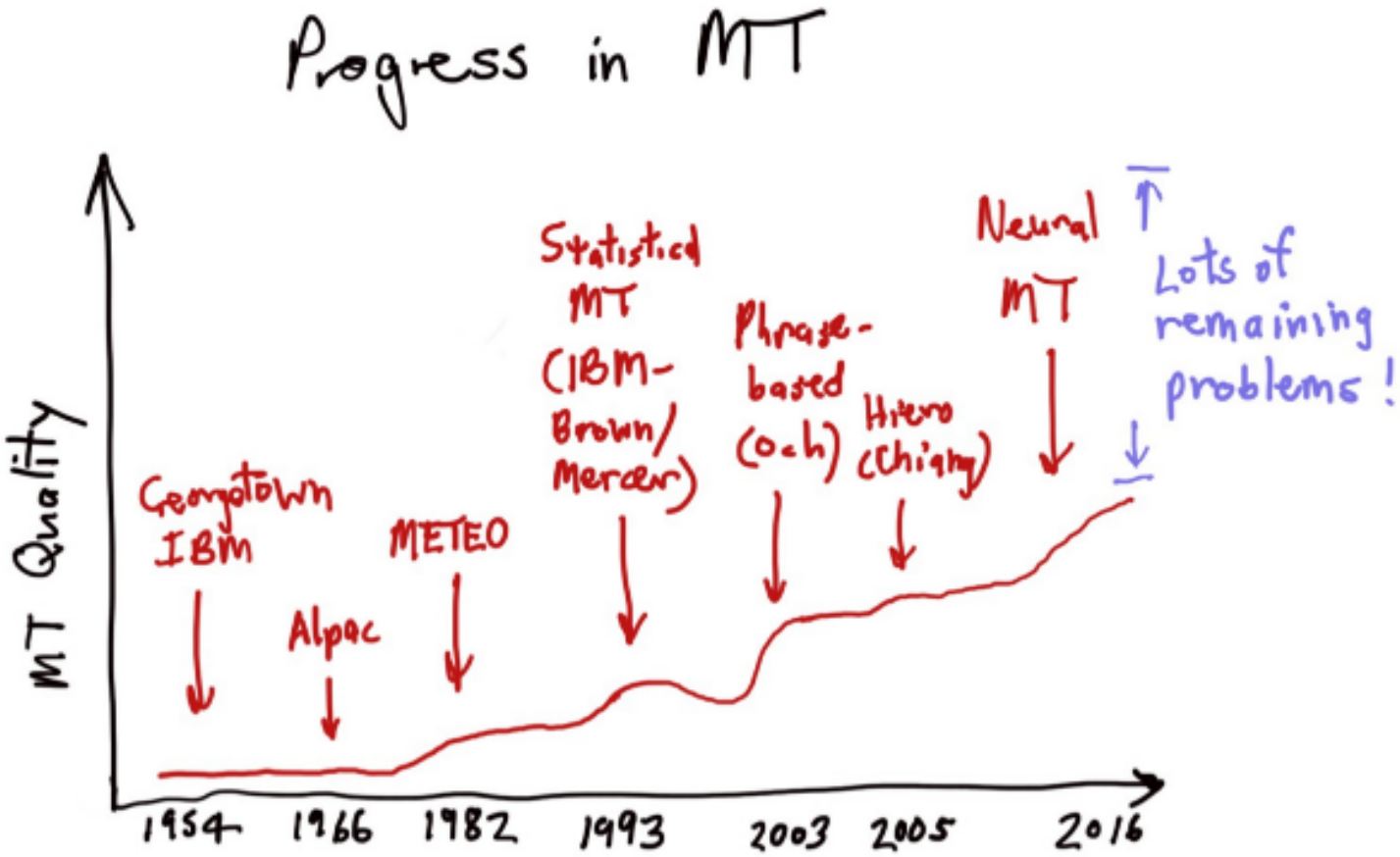
Andy Way and Mary Hearne. 2011. On the Role of Translations in State-of-the-Art Statistical Machine Translation. *Language and Linguistics Compass* **5**:227—248

Now we've gone neural!



Thanks to Kyunghyun Cho

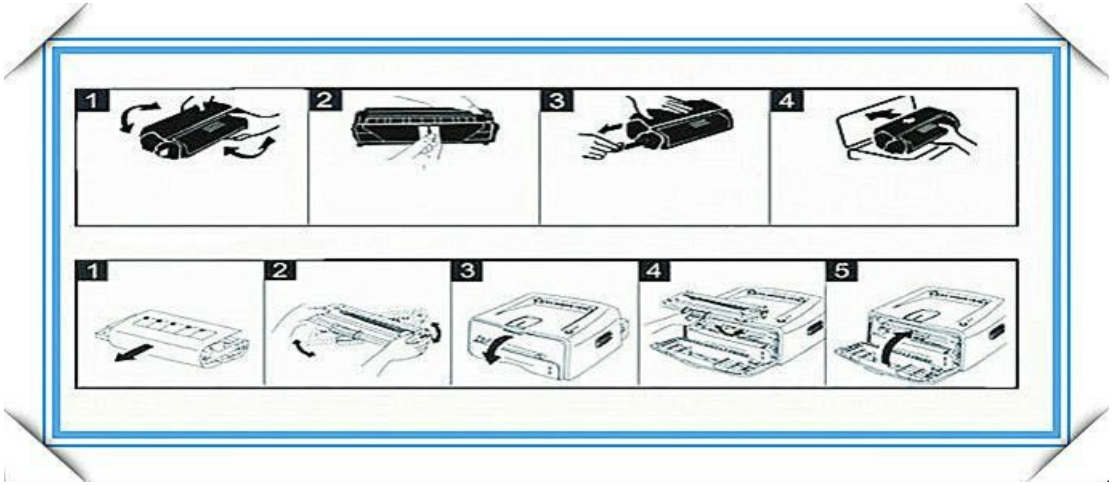
But MT is (still) hard!



Thanks to Chris Manning

MT Evaluation

- Source only!
- Manual:
  - Subjective Sentence Error Rates
  - Correct/Incorrect
  - Error categorization
- Objective Usage Testing



- Automatic MT Evaluation: 100% Match, WER, BLEU, NIST etc.

Human Evaluation of MT

- Fluency and Adequacy scored by subjective human judges
- Inter-rater agreement (and intra-rater agreement!)
- Very expensive (w.r.t time and money)
- Not always available
- Can’t help day-to-day system development
- Imperfect
- …

Automatic Evaluation of MT

- Objective
- Inspired by the WER metric used in ASR
- Measuring the “closeness” between the MT hypothesis and human reference translations
  - Precision: *n*-gram precision
  - Recall:
    - Against the best matched reference
    - Approximated by brevity penalty
- Cheap, fast
- Highly correlated with human evaluations
- MT research has greatly benefited from automatic evaluations
- Typical metrics: BLEU, NIST, F-Score, Meteor, TER ...

BLEU Evaluation Metric

Reference (human) translation:

The US island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself Osama Bin Laden and threatening a biological/chemical attack against the airport.

Machine translation:

The American [?] International airport and its the office a [?] receives one calls self the sand Arab rich business [?] and so on electronic mail, which sends out; The threat will be able after the maintenance at the airport.

- N-gram precision (score between 0 & 1)
  - what % of machine *n*-grams (a sequence of words) can be found in the reference translation?
- Brevity Penalty
  - Can't just type out single word "the" (precision 1.0!)

NB, Extremely hard to game the system, i.e. find a way to change MT output so that BLEU score increases, but quality doesn't (cf. OOV items).

The More Reference Translations, the Better!

Reference translation 1:
The US island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself Osama Bin Laden and threatening a biological/ chemical attack against the airport.

Reference translation 2:
Guam International Airport and its offices are maintaining a high state of alert after receiving an e-mail that was from a person claiming to be the rich Saudi Arabian businessman Osama Bin Laden and that threatened to launch a biological and chemical attack on the airport.

Machine translation:
The American [?] International airport and its the office a [?] receives one calls self the sand Arab rich business [?] and so on electronic mail , which sends out; The threat will be able after the maintenance at the airport to start the biochemistry attack.

Reference translation 3:
The US International Airport of Guam and its office has received an email from a self-claimed Arabian millionaire named Laden , which threatens to launch a biochemical attack on airport. Guam authority has been on alert.

Reference translation 4:
US Guam International Airport and its offices received an email from Mr. Bin Laden and other rich businessmen from Saudi Arabia. They said there would be biochemistry air raid to Guam Airport. Guam needs to be in high precaution about this matter.

BLEU in Action

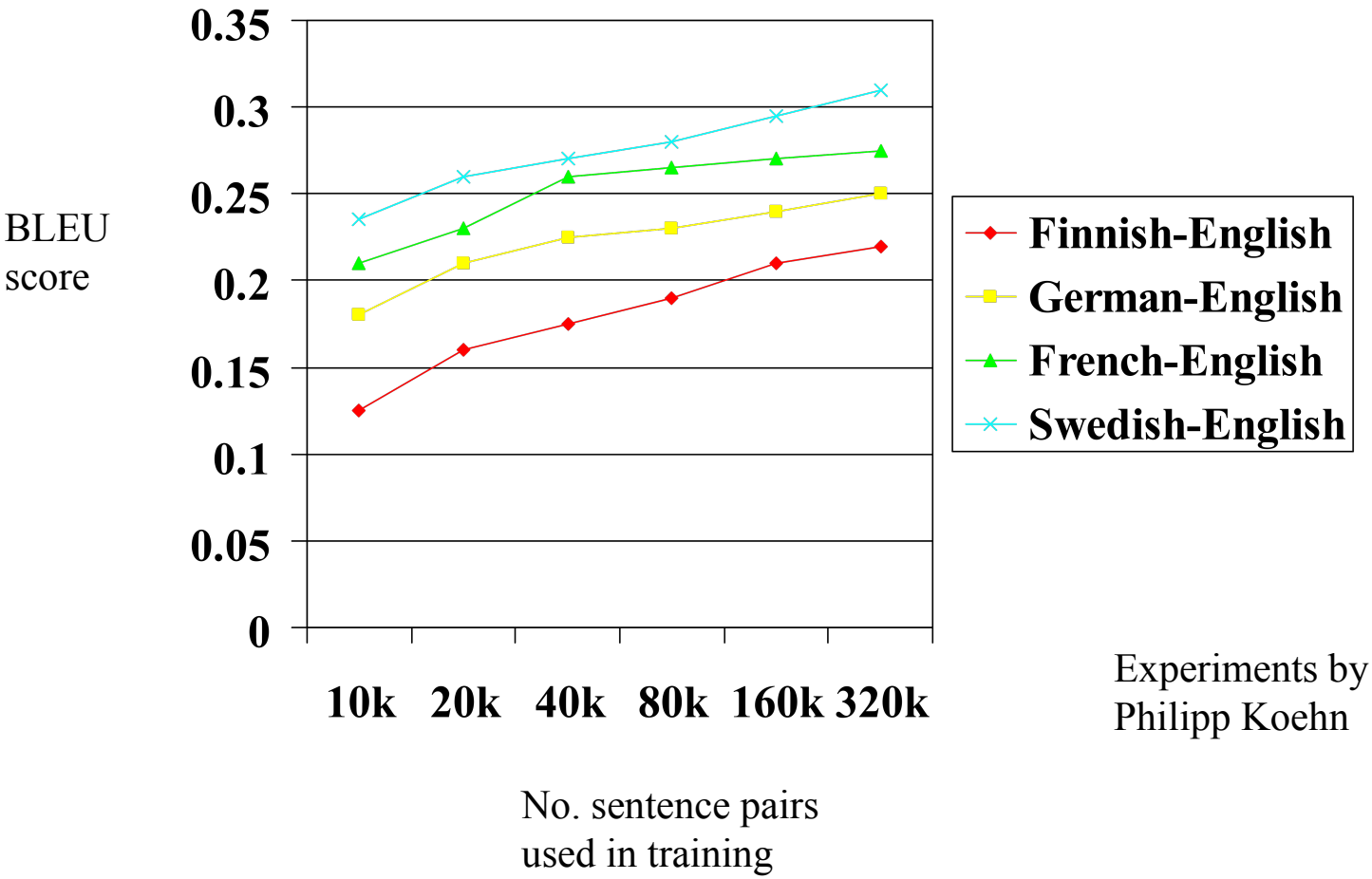
MT Hypothesis: ***the gunman was shot dead by police .***

- Ref 1: The gunman was shot to death by the police .
- Ref 2: The gunman was shot to death by the police .
- Ref 3: Police killed the gunman .
- Ref 4: The gunman was shot dead by the police .

- Precision: p_1=1.0(8/8) p_2=0.86(6/7) p_3=0.67(4/6) p_4=0.6 (3/5)
- Brevity Penalty: c=8, r=9, BP=0.8825
- Final Score:

$$\sqrt[4]{1\times 0.86\times 0.67\times 0.6}\times 0.8825 = 0.68$$

Correlation between BLEU score and Training Set Size



Experiments by Philipp Koehn

Problems with BLEU

1. It can be easy to look good (cf. output from 'state-of-the-art' SMT systems)
2. Not very sensitive to global syntactic structure (disputable …)
3. Doesn't care about nature of untranslated words:
   - *gave it to Bush*
   - *gave it at Bush*
   - *gave it to rhododendron*
4. As MT improves, BLEU won't be 'good enough' …

Problems with *Using* BLEU

- Not designed to test individual sentences
- Not meant to compare different MT systems

- Extremely useful tool for system developers!

Q: what/who is evaluation for?

cf. [Callison-Burch et al., EACL-06; Tiedemann & Hardmeier, EAMT-15]

Other automatic MT metrics

- P&R (GTM: Turian et al., MT-Summit 03)
- RED (Akiba et al., MT-Summit 01) [based on edit distance, cf. WER/PER ...]
- ORANGE (Lin & Och COLING-04)
- Classification by Learning (Kulesza & Shieber TMI-04)
- Meteor (Banerjee & Lavie, ACL-05 ...)
- TER (Snover et al., AMTA 2006)
- ...

Thanks for listening!

The end

Thanks for listening!

~~The end~~

beginning!