20 Years of Statistical Machine Translation

Philipp Koehn, University of Edinburgh

13 May 2009

20 years, roughly

1988 "A Statistical Approach to Language Translation" (Brown et. all, COLING)

2009 a meeting in Prague

Where are we now?

- Comparing rule-based and statistical approaches
- EUROMATRIX organizes yearly evaluation campaign
  - comparing participating research systems
  - benchmarking against off-the-shelf commercial systems
  - task: news translation
- A fair task?
  - translation performance differs across domains, text types, etc.
  - we do not have parallel corpora for news
  - (... we do have monolingual corpora and development sets)
  - off-the-shelf systems had no chance to optimize to task

What works better?

Language Pair	Winner
French-English	statistical
English-French	statistical
German-English	rule-based
English-German	rule-based
Spanish-English	statistical
English-Spanish	tie
Hungarian-English	rule-based
Czech-English	statistical
English-Czech	rule-based

Building statistical systems is quick

	Target Language																				
	en	**bg**	**de**	**cs**	**da**	**el**	**es**	**et**	**fi**	**fr**	**hu**	**it**	**lt**	**lv**	**mt**	**nl**	**pl**	**pt**	**ro**	**sk**	**sl**	**sv**
en	–	40.5	46.8	52.6	50.0	41.0	55.2	34.8	38.6	50.1	37.2	50.4	39.6	43.4	39.8	52.3	49.2	55.0	49.0	44.7	50.7	52.0
bg	61.3	–	38.7	39.4	39.6	34.5	46.9	25.5	26.7	42.4	22.0	43.5	29.3	29.1	25.9	44.9	35.1	45.9	36.8	34.1	34.1	39.9
de	53.6	26.3	–	35.4	43.1	32.8	47.1	26.7	29.5	39.4	27.6	42.7	27.6	30.3	19.8	50.2	30.2	44.1	30.7	29.4	31.4	41.2
cs	58.4	32.0	42.6	–	43.6	34.6	48.9	30.7	30.5	41.6	27.4	44.3	34.5	35.8	26.3	46.5	39.2	45.7	36.5	43.6	41.3	42.9
da	57.6	28.7	44.1	35.7	–	34.3	47.5	27.8	31.6	41.3	24.2	43.8	29.7	32.9	21.1	48.5	34.3	45.4	33.9	33.0	36.2	47.2
el	59.5	32.4	43.1	37.7	44.5	–	54.0	26.5	29.0	48.3	23.7	49.6	29.0	32.6	23.8	48.9	34.2	52.5	37.2	33.1	36.3	43.3
es	60.0	31.1	42.7	37.5	44.4	39.4	–	25.4	28.5	51.3	24.0	51.7	26.8	30.5	24.6	48.8	33.9	57.3	38.1	31.7	33.9	43.7
et	52.0	24.6	37.3	35.2	37.8	28.2	40.4	–	37.7	33.4	30.9	37.0	35.0	36.9	20.5	41.3	32.0	37.8	28.0	30.6	32.9	37.3
fi	49.3	23.2	36.0	32.0	37.9	27.2	39.7	34.9	–	29.5	27.2	36.6	30.5	32.5	19.4	40.6	28.8	37.5	26.5	27.3	28.2	37.6
fr	64.0	34.5	45.1	39.5	47.4	42.8	60.9	26.7	30.0	–	25.5	56.1	28.3	31.9	25.3	51.6	35.7	61.0	43.8	33.1	35.6	45.8
hu	48.0	24.7	34.3	30.0	33.0	25.5	34.1	29.6	29.4	30.7	–	33.5	29.6	31.9	18.1	36.1	29.8	34.2	25.7	25.6	28.2	30.5
it	61.0	32.1	44.3	38.9	45.8	40.6	26.9	25.0	29.7	52.7	24.2	–	29.4	32.6	24.6	50.5	35.2	56.5	39.3	32.5	34.7	44.3
lt	51.8	27.6	33.9	37.0	36.8	26.5	21.1	34.2	32.0	34.4	28.5	36.8	–	40.1	22.2	38.1	31.6	31.6	29.3	31.8	35.3	35.3
lv	54.0	29.1	35.0	37.8	38.5	29.7	8.0	34.2	32.4	35.6	29.3	38.9	38.4	–	23.3	41.5	34.4	39.6	31.0	33.3	37.1	38.0
mt	72.1	32.2	37.2	37.9	38.9	33.7	48.7	26.9	25.8	42.4	22.4	43.7	30.2	33.2	–	44.0	37.1	45.9	38.9	35.8	40.0	41.6
nl	56.9	29.3	46.9	37.0	45.4	35.3	49.7	27.5	29.8	43.4	25.3	44.5	28.6	31.7	22.0	–	32.0	47.7	33.0	30.1	34.6	43.6
pl	60.8	31.5	40.2	44.2	42.1	34.2	46.2	29.2	29.0	40.0	24.5	43.2	33.2	35.6	27.9	44.8	–	44.1	38.2	38.2	39.8	42.1
pt	60.7	31.4	42.9	38.4	42.8	40.2	60.7	26.4	29.2	53.2	23.8	52.8	28.0	31.5	24.8	49.3	34.5	–	39.4	32.1	34.4	43.9
ro	60.8	33.1	38.5	37.8	40.3	35.6	50.4	24.6	26.2	46.5	25.0	44.8	28.4	29.9	28.7	43.0	35.8	48.5	–	31.5	35.1	39.4
sk	60.8	32.6	39.4	48.1	41.0	33.3	46.2	29.8	28.4	39.4	27.4	41.8	33.8	36.7	28.5	44.4	39.0	43.3	35.3	–	42.6	41.8
sl	61.0	33.1	37.9	43.5	42.6	34.0	47.0	31.1	28.8	38.2	25.7	42.3	34.6	37.3	30.0	45.9	38.2	44.1	35.8	38.9	–	42.7
sv	58.5	26.9	41.0	35.6	46.6	33.3	46.6	27.4	30.9	38.9	22.7	42.0	28.2	31.0	23.7	45.6	32.2	44.2	32.7	31.3	33.5	–

462 translation systems for all but one official EU-27 languages, using Acquis corpus

Why are some language pairs harder?

- Simple linear regression models showing correlation of BLEU with explanatory factors. Extension of Birch et al. [EMNLP 2008]

Factor	R^2	**Significant?**
Phrase translation entropy	0.276	***
Reordering amount	0.267	***
Language relatedness	0.115	***
Target vocabulary size	0.101	***
Source corpus size	0.034	***
Target corpus size	0.034	***
Source vocabulary size	0.001	

- These factors explain 74.5% of score differences

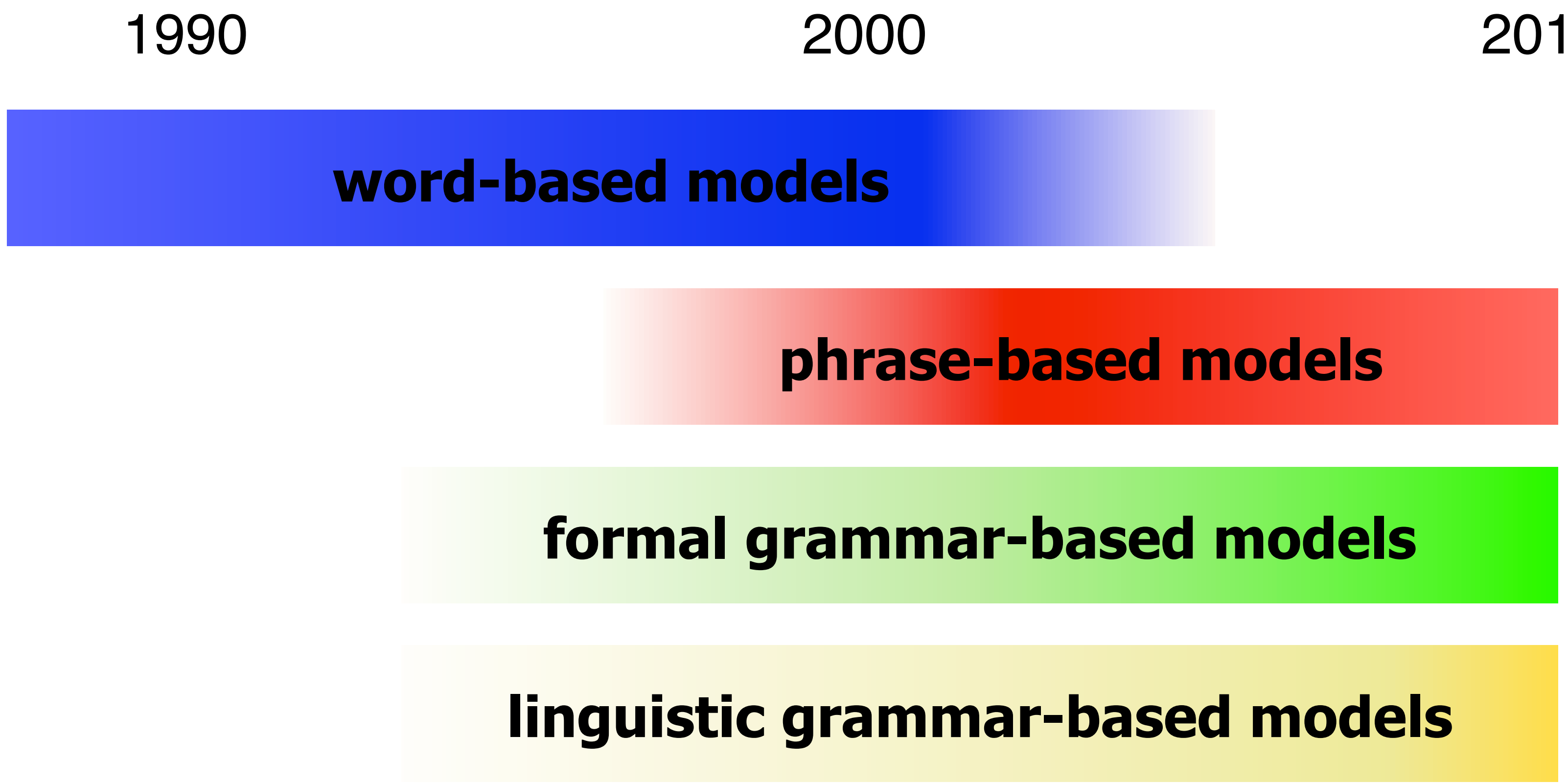
Where are we going?

Linguistics

Machine Learning

Human-Computer Interaction

Linguistics: Progress



Translation Rules

- Phrase translation

$$\text{the house} \rightarrow \text{das Haus}$$

- Factored phrase translation

$$\binom{\text{the}}{\mathsf{det}} \binom{\text{house}}{\mathsf{n}} \rightarrow \binom{\text{das}}{\mathsf{det}} \binom{\text{Haus}}{\mathsf{n}}$$

- Hierarchical phrase translation [Chiang, ACL 2005]

$$\text{must seek } \textsc{x} \rightarrow \text{muss } \textsc{x} \text{ suchen}$$

- Syntactified translation [Marcu et al., ACL 2006]

$$\textsc{s}\ [\ \textsc{np}_1 \text{ must seek } \textsc{np}_2\ ] \rightarrow \textsc{s}\ [\ \textsc{np}_1 \text{ muss } \textsc{np}_2 \text{ suchen}\ ]$$

The Future: Syntax

- Phrase-structure grammar or dependency structure?
- Context-sensitive, context-free?
- Syntax at the source or the target?
- Automatically learn transfer syntax, or use tree-banks, rules?
- S-CFG, S-TIG, S-TAG, CCG, LFG, ... ?

Lexical Semantics

- Words have different meanings, we need to distinguish them.

- bank

  1. financial institution
  2. shore of a river

- Statistical machine translation already handles this rather well.

Lexical Semantics

- Statistical models use context words as features to determine word sense

- money and depost indicate financial sense

  After collecting the money, he went to deposit it in the bank.

- sand and ships indicate river sense

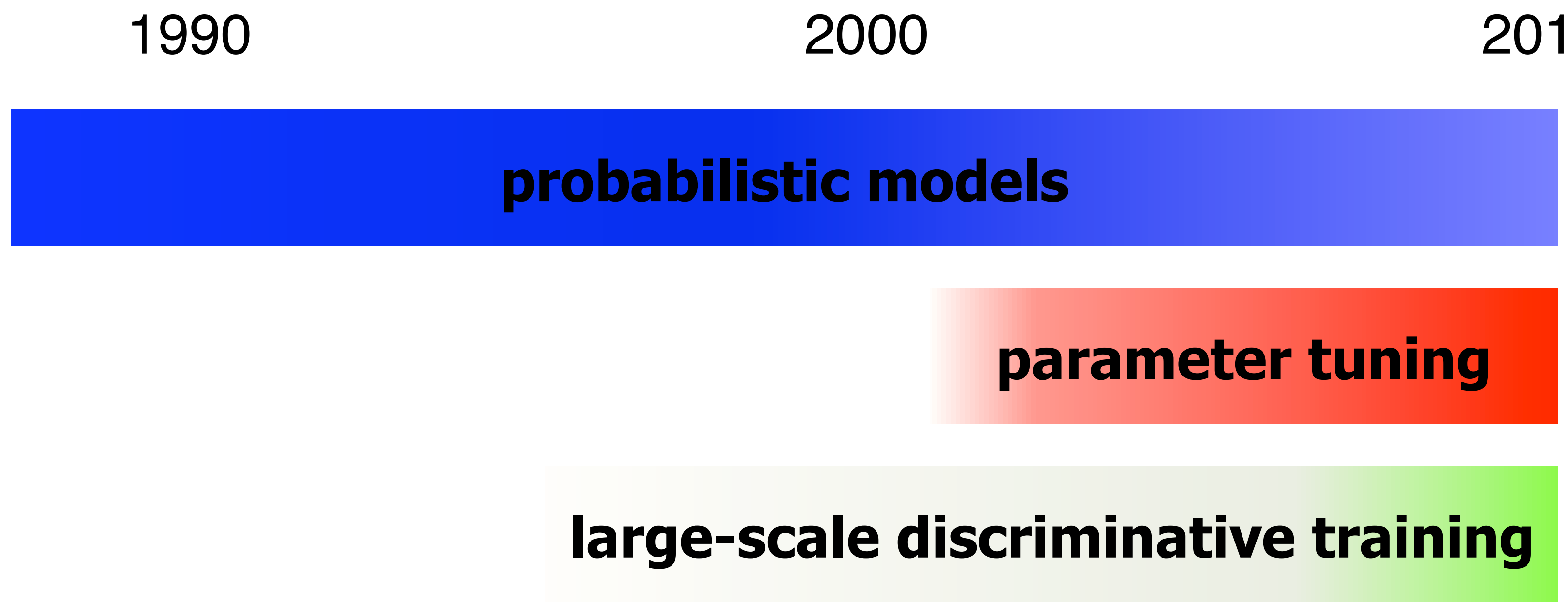
  She sat in the sand at the bank, gazing at the ships in the distance.

Inference Semantics

He was more comfortable with his female relatives. He did not like his brothers, but he loved spending the summer with his cousin.

- When translating cousin into English, you need to determine the gender.
- Required:
  - anaphora resolution that relatives and cousin co-refer
  - inference that comfortable with and loved spending are connected
- We have made little progress on this.

Machine Learning: Progress



Machine Learning Methods

- There are many parameter in a statistical machine translation system
  - language model n-grams
  - translation rules
  - reordering features
  - syntactic relationships
  - impact of context features
  - relative importance of language model and translation model

- Should we model the training data or optimize on translation performance?

$$\text{maximize } p(\text{DATA}) \Leftrightarrow \text{maximize BLEU}$$

- The big problem: scaling to millions of features, millions of sentence pairs

Data

Don't think about algorithms, get more data.
If you want to think, think about getting more data.
Eric Brill, 2001

- Getting more data
  - crawling the web for parallel corpora
  - acquiring translation memories from language service providers
- Thinking about getting more data
  - collaborating with users - WikiTrans

Human-Computer Interaction

- Main application of machine translation: gisting
- But: much bigger need for publication-quality translation
- How can machine translation help human translators?
  - translation memories are industry standard
  - post-editing machine translation used increasingly
- Better interactions?

Post-Editing

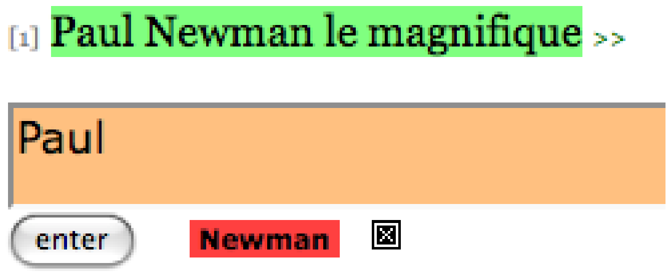
<< [2] L'inoubliable interprète de "Butch Cassidy et le Kid" est mort des suites d'un cancer, à l'âge de 83 ans, dans sa maison du Connecticut. >>

The unforgettable ~~interpreter~~ actor of " Butch Cassidy and the Sundance Kid " died as a result of cancer ~~,~~ at the age of 83 ~~years~~ ~~,~~ in his house in Connecticut . (9 edits)

The unforgettable actor of "Butch Cassidy and the Sundance Kid" died as a result of cancer at the age of 83 in his house in Connecticut.

- Correcting machine translation faster than translating from scratch?
  - faster and better: yes
  - more enjoyable: no

Trans-Type: Sentence Completion



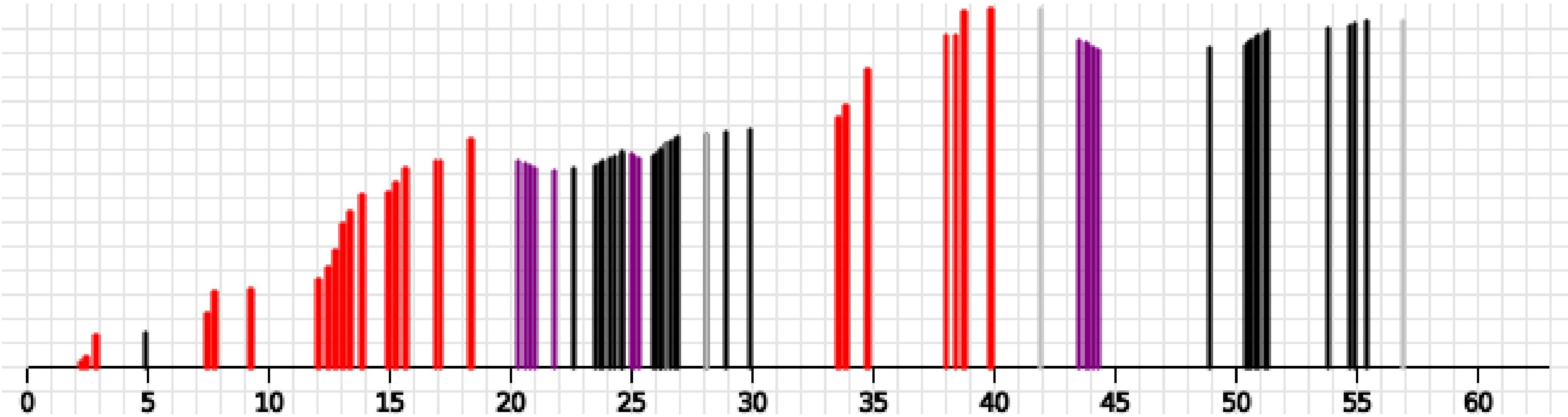
- Based on work of the EC project TransType2
  - system makes suggestion how to complete the sentence
  - user accepts it, or types in own translation
  - system computes new suggestion

Other Types of Assistance

Paul	Newman	le	magnifique
Paul	**Newman**	**the**	**wonderful**
Mr	Newman ,	**the magnificent**	
Mr Paul	Newman here	the wonderful	
as Paul	Committee		beautiful
another	Newman , who speaks		magnificent
with Paul		the splendid	
, Paul		the excellent	
of Paul		the beautiful	
work of Paul			it
the words of Paul			great

- Translation options from the phrase table, ranked by probability
- Many other types of assistance possible (confidence, fluency models, ...)

Logging the Activity



red: accepting translation predcition, black: keystroke, purple: deletion, grey: cursor movement
x-axis: time in seconds, y-axis: length of translation in characters

- Enables insight into the translation process
- Helps with improving translation tools

Our Translation Tool

- Work carried out within EuroMatrixPlus
- Available online: `http://tool.statmt.org/`
- User study [MT Summit, 2009, submitted]
  - users faster and better with each type of assistance
  - but: better translators often ignore assistance
  - fastest and best with post-editing, but self-report that it is less useful

Final Comments

- A vibrant field
  - rapid progress, fueled by competitions
  - new ideas spread quickly
- Progress on many fronts
  - linguistics
  - machine learning
  - tools for translators
- Engaging the community
  - open source tools and corpora
  - many stake-holders, many languages