

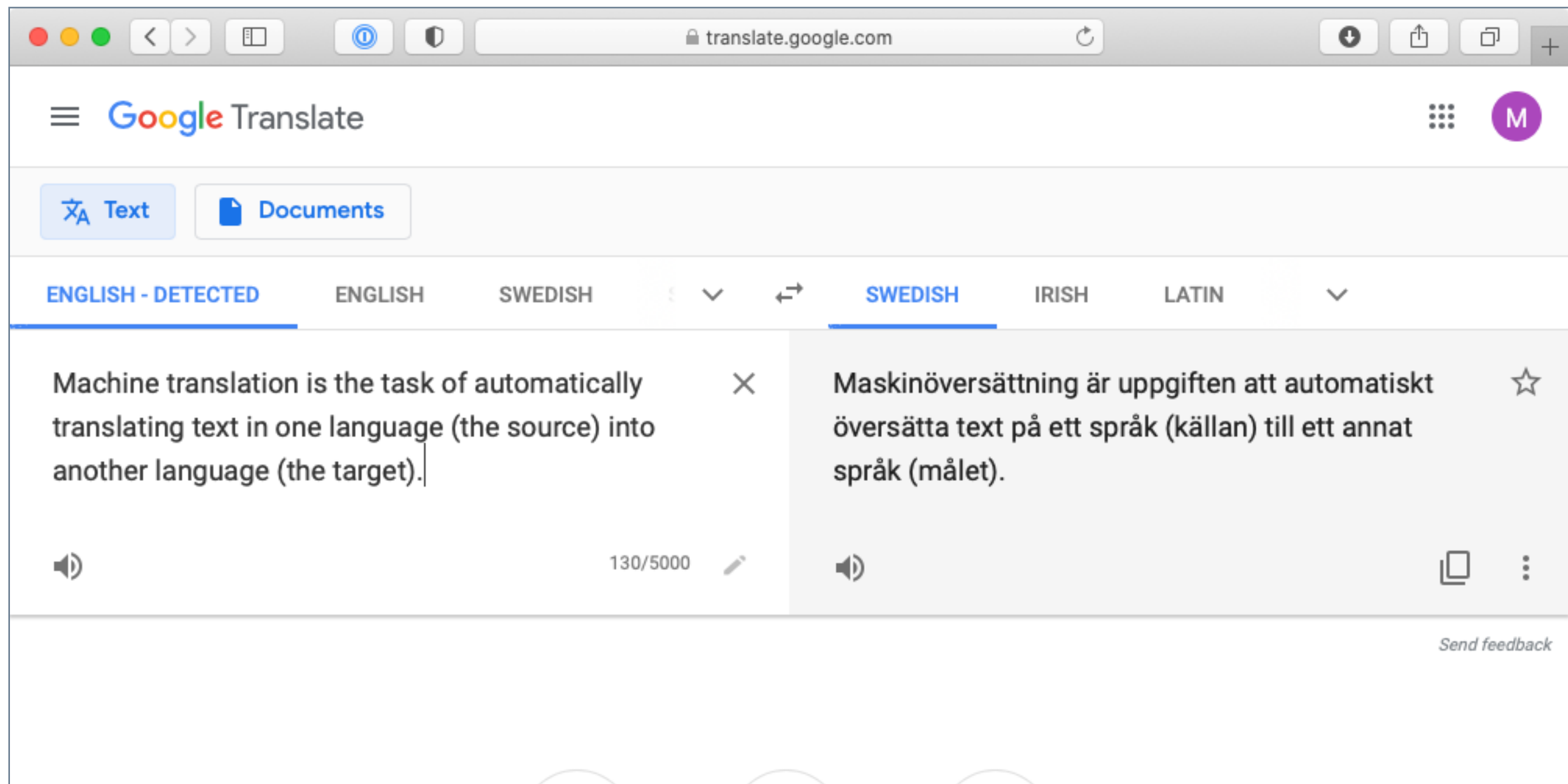
Natural Language Processing

Introduction to machine translation

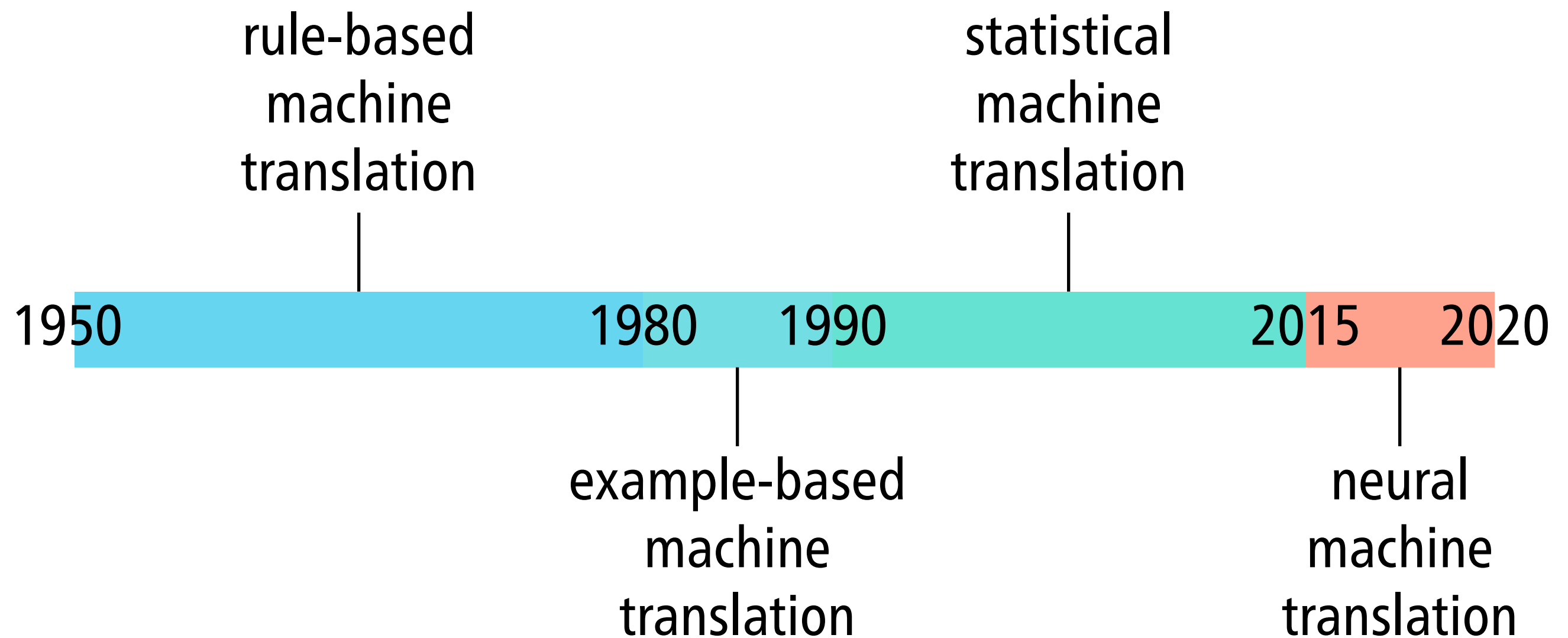
Marco Kuhlmann

Department of Computer and Information Science

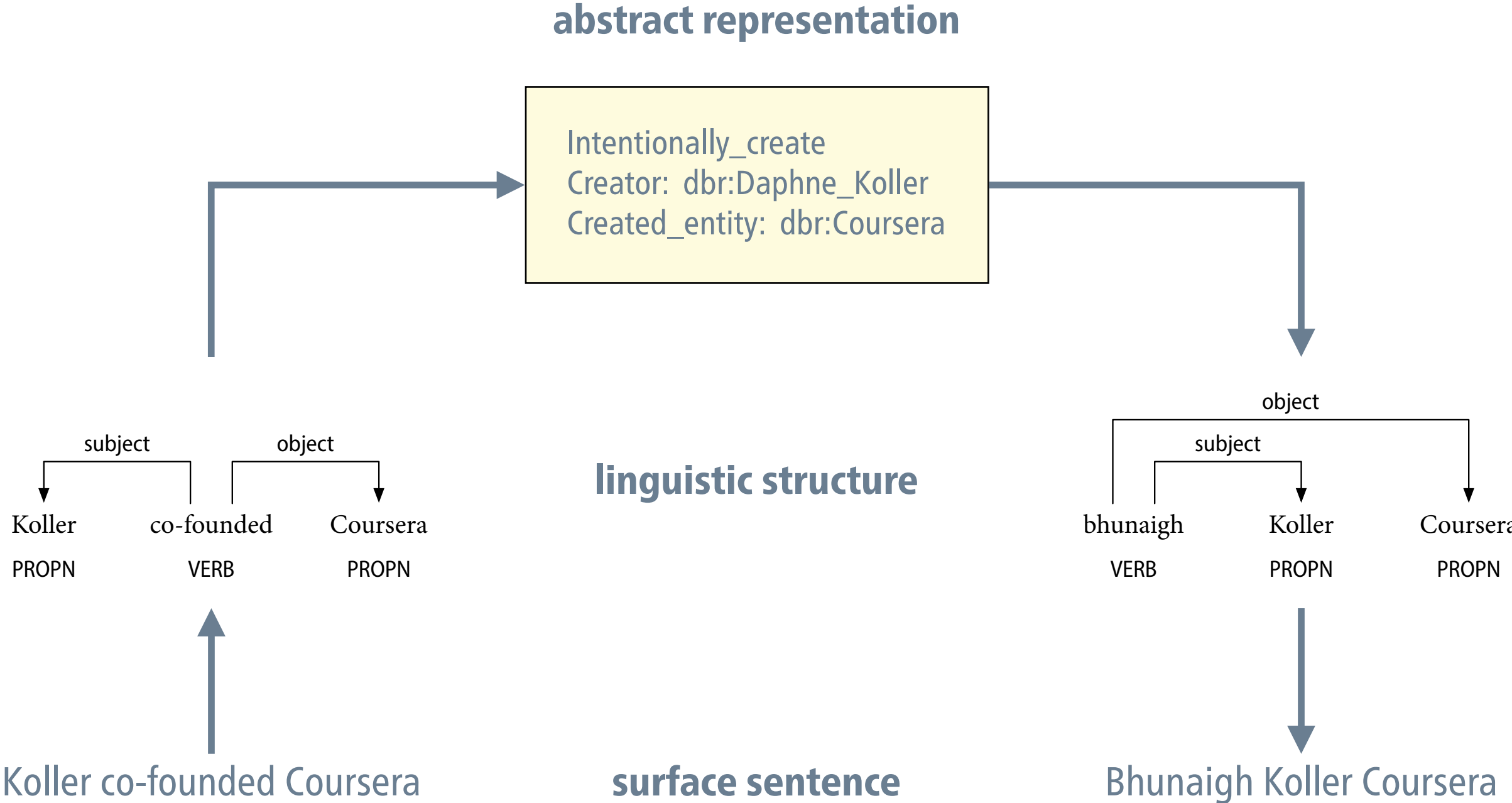
Machine translation



A timeline of machine translation



Interlingual machine translation



Noisy Channel Model

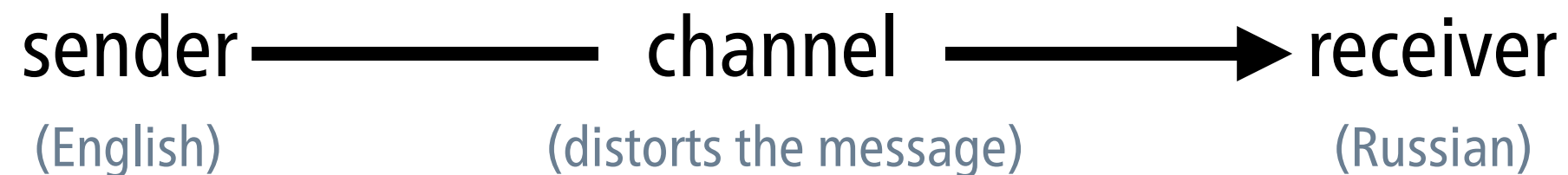


Image source



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)

Statistical machine translation (SMT)

- Formulate machine translation as an optimisation task: Given a source sentence $\mathbf{x}$, find the most probable target sentence $\mathbf{y}$:

$$\arg \max_{\mathbf{y}} P(\mathbf{y} | \mathbf{x})$$

- Use Bayes' rule to decompose the probability model into two components that can be learned separately:

The diagram illustrates the decomposition of the SMT equation $\arg \max_{\mathbf{y}} P(\mathbf{x} | \mathbf{y}) P(\mathbf{y})$ into its constituent parts. A horizontal line connects the text 'decoding algorithm heuristic search' to the equation. From the equation, two vertical lines descend to the text 'translation model learnt from parallel data' and 'language model learnt from monolingual data'.

decoding algorithm
heuristic search

$\arg \max_{\mathbf{y}} P(\mathbf{x} | \mathbf{y}) P(\mathbf{y})$

translation model
learnt from parallel data

language model
learnt from monolingual data

Parallel corpora

- Canadian Hansard (English–French); extracted from the proceedings of the Canadian Parliament.
- Europarl (21 languages); 30.32 M parallel sentences extracted from the proceedings of the European parliament.

[Link to the Europarl website](#)

- OPUS (several languages); growing collection of translated texts, automatically preprocessed and aligned.

[Link to the OPUS website](#)

Word-to-word alignments

	the	house	is	small
das				
haus				
ist				
klein				

	the	house	is	small
klein				
ist				
das				
haus				

	the	house	is	very	small
das					
haus					
ist					
klitzeklein					

	huset	är	litet
das			
haus			
ist			
klein			

	the	house	is	just	small
das					
haus					
ist					
klein					

	er	hat	ins	gras	gebissen
he					
kicked					
the					
bucket					

IBM Model 1

$$P(\underset{\substack{| \\ \text{alignment function}}}{\boldsymbol{x}}, \boldsymbol{a} \mid \boldsymbol{y}) = \frac{\overset{\substack{\text{normalisation constant} \\ | \\ \varepsilon}}{\varepsilon}}{(|\boldsymbol{y}| + 1)^{|\boldsymbol{x}|}} \prod_{j=1}^{|\boldsymbol{x}|} t(x_j \mid y_{a(j)})$$

- First in a series of increasingly complex statistical translation models; deals only with lexical (word-to-word) translation.
- Central component: The lexical translation probability t of observing a source word x , given the aligned target word y .

Training statistical machine translation models

- We would like to estimate the lexical translation probabilities from a parallel corpus – but we do not have the alignments.
- We can bootstrap the translation probabilities and alignments in parallel using the Expectation–Maximization (EM) algorithm:
 1. initialise the model parameters randomly
 2. calculate alignments based on the current model parameters
 3. estimate new model parameters from the new alignments
 4. repeat steps 2–3 until convergence

[Brown et al. \(1993\)](#)

Statistical machine translation (SMT)

- Research on statistical machine translation led to significant improvements in the availability and quality of translation.

The first version of Google Translate (2006–2016) was an SMT system.

- However, the best systems were extremely complex and required large amounts of external resources and feature engineering.

Open-source example: Moses

Evaluation: BLEU (Bilingual Evaluation Understudy)

- BLEU compares the automatic translation of a source sentence to one or several human-created reference translations.
- BLEU combines n -gram precision (for n up to 4) with a brevity penalty for too-short translations.
- BLEU has been criticised for not correlating well with human judgement, and several other evaluation measures exist.

[Papineni et al. \(2002\); Callison-Burch et al. \(2006\)](#)