

Počítačová lingvistika a strojový překlad

Zcela nový svět díky hlubokému učení

Ondřej Bojar

bojar@ufal.mff.cuni.cz

Ústav formální a aplikované lingvistiky

Matematicko-fyzikální fakulta

Univerzita Karlova

3. prosinec 2018

Osnova

- ▶ Počítačová lingvistika a její problémy a aplikace.

Strojový překlad:

- ▶ Proč je překlad těžký
- ▶ Předneuronový strojový překlad (SMT).
- ▶ (Mezihra: hluboké strojové učení.)
- ▶ Neuronový strojový překlad (NMT).
 - ▶ Aktuální kvalita překladu.
- ▶ Lidské a umělé reprezentace významu.

Na hranici oborů...

čeština,
angličtina, němčina ...

matematika

počítače

Počítačová lingvistika



Lingvistické problémy

- ▶ Víceznačnost a význam slov.

Spal celou Petkevičovu přednášku. Ženu holí stroj.

- ▶ Bohatost slovních tvarů.

- ▶ Čeština má 7 pádů, 3 čísla a 4 rody.
- ▶ V angličtině se používá ~50 morfologických značek, v češtině 4000.

- ▶ Koordinace a apozice:

Předseda vlády, Petr Nečas[□] a Martin Lhota namítli...

Anglicky We have both countries inside and outside the Eurozone.

Význam 1 Máme tu země eurozóny a země stojící mimo eurozónu.

Význam 2 Máme obě země uvnitř a vně eurozóny.

- ▶ Zájmena.

Ukázka bohatosti jazyka

Kolik je správných překladů následující věty?

And even though he is a political veteran, the Councilor
Karel Brezina responded similarly.

Ukázka bohatosti jazyka

Příklady ze 71 tisíc správných překladů anglické věty:

And even though he is a political veteran, the Councilor
Karel Brezina responded similarly.

A ačkoli ho lze považovat za politického veterána, radní Brezina reagoval obdobně.

Ač ho můžeme prohlásit za politického veterána, reakce radního Karla Breziny byla v

A i přestože je politický matador, radní Karel Brezina odpověděl podobně.

A přestože je to politický veterán, velmi obdobná byla i reakce radního K. Breziny.

A radní K. Brezina odpověděl obdobně, jakkoli je politický veterán.

A třebaže ho můžeme považovat za politického veterána, reakce Karla Breziny byla v

Byť ho lze označit za politického veterána, Karel Brezina reagoval podobně.

Byť ho můžeme prohlásit za politického veterána, byla i odpověď K. Breziny velmi po

K. Brezina, i když ho lze prohlásit za politického veterána, odpověděl velmi obdobně.

Odpověď Karla Breziny byla podobná, navzdory tomu, že je politickým veteránem.

Radní Brezina odpověděl velmi obdobně, navzdory tomu, že ho lze prohlásit za politického

Reakce K. Breziny, třebaže je politický veterán, byla velmi obdobná.

Velmi obdobná byla i odpověď Karla Breziny, ačkoli ho lze prohlásit za politického ve

Lingvistické nástroje

- ▶ Identifikace kódování dokumentu a jazyka.
- ▶ Rozpoznání hranic vět a slov:
Švejk 12. prosince dorazil na král. Vinohrady s dopisem.
ajskrím → I scream / icecream.
- ▶ Morfologická analýza.
- ▶ Povrchový a hloubkový větný rozbor.
- ▶ Identifikace pojmenovaných entit:
Bílý dům se nechal slyšet.
Rice University ≠ univerzita rýže
- ▶ Koreference (mj. identifikace, co zastupují zájmena).

Lingvistická data

- ▶ **Korpusy** jsou (velké) sbírky textů:
 - ▶ Texty typicky označované nebo včetně větných rozborů.
Pražský závislostní korpus (PDT): 1.5 mil. slov.
Pražský čj-aj závislostní korpus (PCEDT): 50 tis. vět.
 - ▶ Některé vícejazyčné: CzEng (50 mil. dvojic vět, odpovídá ~200 metrům knih).
- ▶ **Slovníky** na ÚFALu jsou strojově čitelné:
 - ▶ *Morfologický slovník* říká, že *kočka* je české slovo a *kočke* ne.
 - ▶ *Valenční slovník* říká, že:
 - Rodiče přijali Petra.* → je správně
 - Rodiče přijeli Petra.* → není správně
 - ▶ *Slovník subjektivity* obsahuje hodnotící výrazy.

Lingvistické aplikace

	Děláme	
	Dobře	Aktuálně
Kontrola překlepů	**	*
Kontrola pravopisu	**	
Vyhledávání dokumentů	*	
Dolování informací z textu	*	
Automatická sumarizace textů	partneři	**
Syntéza a rozpoznávání mluvené řeči	*	**
Dialogové systémy (chatboti)	*	
Strojový překlad	**	***
Strojový překlad mluvené řeči	*	**
Analýza smýšlení (sentiment)	*	**

Strojový překlad

Machine Translation

Evropská unie: 24 jazyků



- ▶ DGT (Ředitelství překladů):
 - ▶ 1750 lingvistů, 600 asistentů
 - ▶ Miliony stran ročně: 0,9 (1992) 1,2 (2004) 1,8 (2012)
- ▶ Celá EU (2013):
 - ▶ ~6000 překladatelů, 1 miliarda EUR ročně.

MT Must Guess Word Meaning...

I	saw	two	green	striped	cats	.
já	pila	dva	zelený	pruhovaný	kočky	.
	pily	dvě	zelená	pruhovaná	koček	
	...	dvou	zelené	pruhované	kočkám	
	viděl	dvěma	zelení	pruhovaní	kočkách	
	viděla	dvěmi	zeleného	pruhovaného	kočkami	
	...		zelených	pruhovaných		
	uviděl		zelenému	pruhovanému		
	uviděla		zeleným	pruhovaným		
	...		zelenou	pruhovanou		
	viděl jsem		zelenými	pruhovanými		
	viděla jsem			

... and Pick the Right Translation ...

I	saw	two	green	striped	cats	.
já	pila	dva	zelený	pruhovaný	kočky	.
	pily	dvě	zelená	pruhovaná	koček	
	...	dvou	zelené	pruhované	kočkám	
	viděl	dvěma	zelení	pruhovaní	kočkách	
	viděla	dvěmi	zeleného	pruhovaného	kočkami	
	...		zelených	pruhovaných		
	uviděl		zelenému	pruhovanému		
	uviděla		zeleným	pruhovaným		
	...		zelenou	pruhovanou		
	viděl jsem		zelenými	pruhovanými		
	viděla jsem			

... But Context Matters

I	saw	two	green	striped	cats	.
já	pila	dva	zelený	pruhovaný	kočky	.
	pily	dvě	zelená	pruhovaná	koček	
	...	dvou	zelené	pruhované	kočkám	
	viděl	dvěma	zelení	pruhovaní	kočkách	
	viděla	dvěmi	zeleného	pruhovaného	kočkami	
	...		zelených	pruhovaných		
			zelenému	pruhovanému		
zrak mi utkvěl na			zeleným	pruhovaným		
			zelenou	pruhovanou		
			zelenými	pruhovanými		
viděl jsem				
viděla jsem						

Non-Neural Statistical MT

Given a source (foreign) language sentence $f_1^l = f_1 \dots f_j \dots f_J$,
Produce a target language (English) sentence

$e_1^l = e_1 \dots e_j \dots e_I$.

Among all possible e_1^l , choose the most likely one:

$$\hat{e}_1^l = \operatorname{argmax}_{l, e_1^l} p(e_1^l | f_1^l) \quad (1)$$

Use Bayes' law to divide the model into components:

$$\hat{e}_1^l = \operatorname{argmax}_{l, e_1^l} p(f_1^l | e_1^l) p(e_1^l) \quad (2)$$

$p(f_1^l | e_1^l)$ Translation model (TM, "reversed", $e_1^l \rightarrow f_1^l$)
 $p(e_1^l)$ Language model (LM)

Brute-Force MT (1/2)

Only sentences from a “translation memory” (TM) known:

Good morning. = Dobré ráno.
How are you? = Jak se máš?
How are you? = Jak se máte?

$$p(e_1^l | f_1^l) = \begin{cases} 1 & \text{if } e_1^l = f_1^l \text{ seen in the TM} \\ 0 & \text{otherwise} \end{cases}$$

Any problems with the definition?

Brute-Force MT (2/2)

Only sentences from a “translation memory” (TM) known:

Good morning. = Dobré ráno.
How are you? = Jak se máš?
How are you? = Jak se máte?

$$p(e_1^l | f_1^l) = \begin{cases} 1 & \text{if } e_1^l = f_1^l \text{ seen in the TM} \\ 0 & \text{otherwise} \end{cases}$$

- ▶ Not a probability. There may be f_1^l , s.t. $\sum_{e_1^l} p(e_1^l | f_1^l) > 1$.
⇒ Have to normalize, use $\frac{\text{count}(e_1^l, f_1^l)}{\text{count}(f_1^l)}$ instead of 1.
- ▶ Not “smooth”, no generalization:

Good morning. ⇒ Dobré ráno.
Good evening. ⇒ ∅

Phrase-Based Statistical MT (PBMT)

					Total	Weight	Weighted	
Phrase log. prob.	0,0	-0,69	-1,39		-2,08	2,0	-4,16	
Phrase penalty	1,0	1,0	1,0		3,0	-1,0	-3,0	
Word penalty	1,0	2,0	1,0		4,0	-0,5	-2,0	
	Peter	left for	home .					
	▷	Petr	odešel	domů .	◁			
Bigram log. prob.	-4,02	-2,50	-3,61	-0,39	-0,08	-10,59	1,0	-10,59
							Total	-19,75

- ▶ Output composed of selected translation units (“minimum translation units”, MTUs: here phrase pairs, solid)
- ▶ Scored with LM n -grams (dashed) and other features.
- ▶ Scores weighted (log-linear, not pure Bayes actually).

1: PBMT Illustrated

Nemám žádného psa.

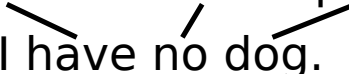
I have no dog.

Viděl kočku.


He saw a cat.

2: Align Words

Nemám žádného psa.
I have no dog.



Viděl kočku.
He saw a cat.



3: Extract Phrase Pairs (MTUs)

Nemám žádného psa.
I have no dog.

Viděl kočku.
He saw a cat.

4: New Input

Nemám žádného psa.
I have no dog.

Viděl kočku.
He saw a cat.

New input: Nemám kočku.

4: New Input

Nemám žádného psa.
I have no dog.

Viděl kočku.
He saw a cat.

... I don't have cat.

New input: Nemám kočku.

5: Pick Probable Phrase Pairs (TM)

Nemám žádného psa.
I have no dog.

Viděl kočku.
He saw a cat.

... I don't have cat.

New input: Nemám kočku.
I have

6: So That n -Grams Probable (LM)

Nemám žádného psa.
I have no dog.

Viděl kočku.
He saw a cat.

... I don't have cat.

New input:

Nemám kočku.
I have a cat.

Meaning Got Reversed!

Nemám žádného psa.
I have no dog.

Viděl kočku.
He saw a cat.

... I don't have cat.

New input:

Nemám kočku.
I have a cat.



What Went Wrong?



$$\hat{e}_1^l = \operatorname{argmax}_{l, e_1^l} p(f_1^l | e_1^l) p(e_1^l) = \operatorname{argmax}_{l, e_1^l} \prod_{(f, \hat{e}) \in \text{phrase pairs of } f_1^l, e_1^l} p(\hat{f} | \hat{e}) p(e_1^l) \quad (3)$$

- ▶ Too strong phrase-independence assumption.
 - ▶ Phrases do depend on each other.
Here “nemám” and “žádného” jointly express one negation.
 - ▶ Word alignments ignored that dependence.
But adding it would increase data sparseness.
- ▶ Language model is separate from translation model.
 - ▶ $p(e_1^l)$ models the target sentence independently of f_1^l .

Redefining $p(e'_1 | f'_1)$

What if we modelled $p(e'_1 | f'_1)$ directly, word by word:

$$\begin{aligned} p(e'_1 | f'_1) &= p(e_1, e_2, \dots, e_l | f'_1) \\ &= p(e_1 | f'_1) \cdot p(e_2 | e_1, f'_1) \cdot p(e_3 | e_2, e_1, f'_1) \dots \\ &= \prod_{i=1}^l p(e_i | e_1, \dots, e_{i-1}, f'_1) \end{aligned} \tag{4}$$

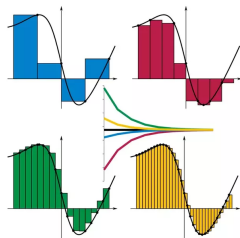
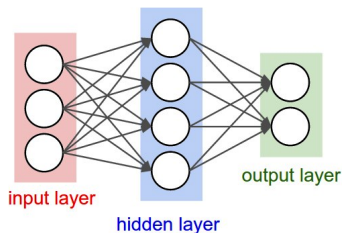
...this is “just a cleverer language model:” $p(e'_1) = \prod_{i=1}^l p(e_i | e_1, \dots, e_{i-1})$

Main Benefit: All dependencies available.

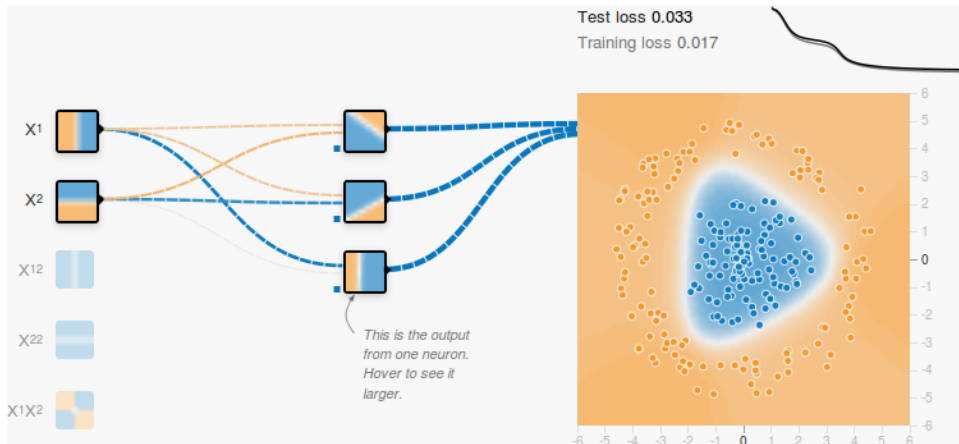
But what technical device can learn this?

Interlude:
Neural Networks
and Representation Learning

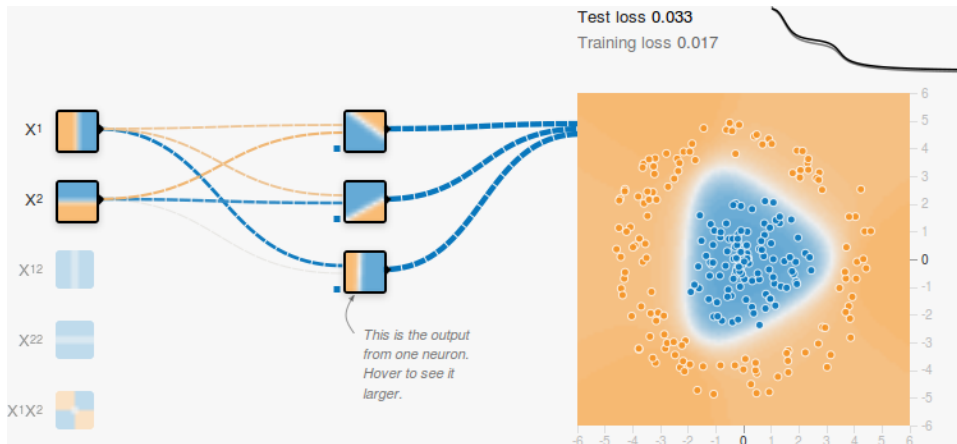
NNs: Universal Approximators



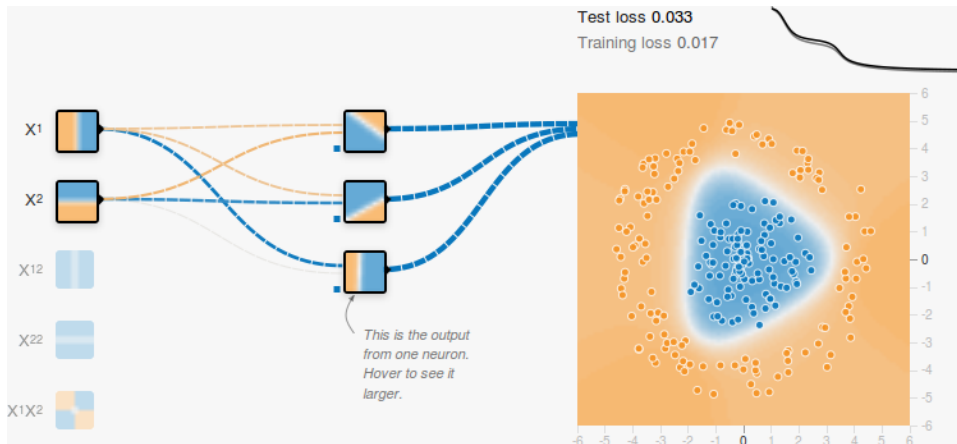
- ▶ A neural network with a single hidden layer (possibly huge) can approximate any continuous function to any precision.
 - ▶ (Nothing claimed about learnability.)
 - ▶ There exists a two-layer neural network (ReLU activations) and $2n + d$ weights that can represent any function on a sample of size n in d dimensions. (Zhang et al., 2016)
- ⇒ Big risks of overfitting.



$$\begin{aligned} & -0.43x_1 - 0.89x_2 + 2.0 > 0 \\ \text{and } & -0.67x_1 + 0.89x_2 + 2.1 > 0 \\ \text{and } & 1.4x_1 - 0.067x_2 + 2.3 > 0 \end{aligned}$$

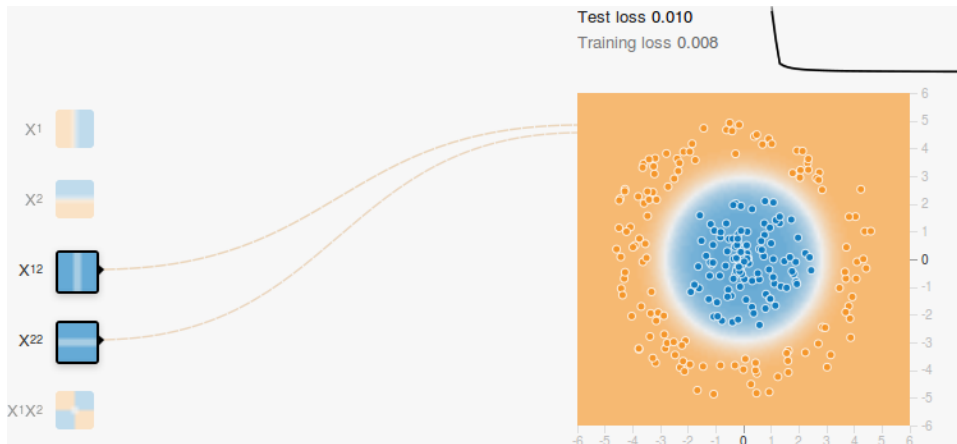


$$\begin{aligned} \text{In fact: } & 1 \tanh(-0.43x_1 - 0.89x_2 + 2.0) \\ & + 1 \tanh(-0.67x_1 + 0.89x_2 + 2.1) \\ & + 1 \tanh(1.4x_1 - 0.067x_2 + 2.3) - \pi/2 > 0 \end{aligned}$$



$$\begin{aligned} \text{In fact: } & 1 \tanh(-0.43x_1 - 0.89x_2 + 2.0) \\ & + 1 \tanh(-0.67x_1 + 0.89x_2 + 2.1) \\ & + 1 \tanh(1.4x_1 - 0.067x_2 + 2.3) - \pi/2 > 0 \end{aligned}$$

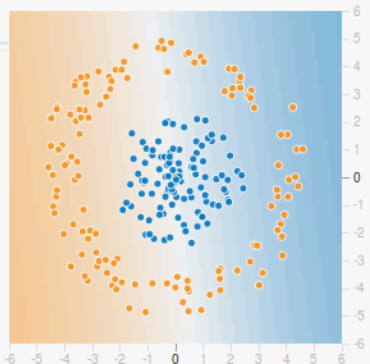
Perfect Features



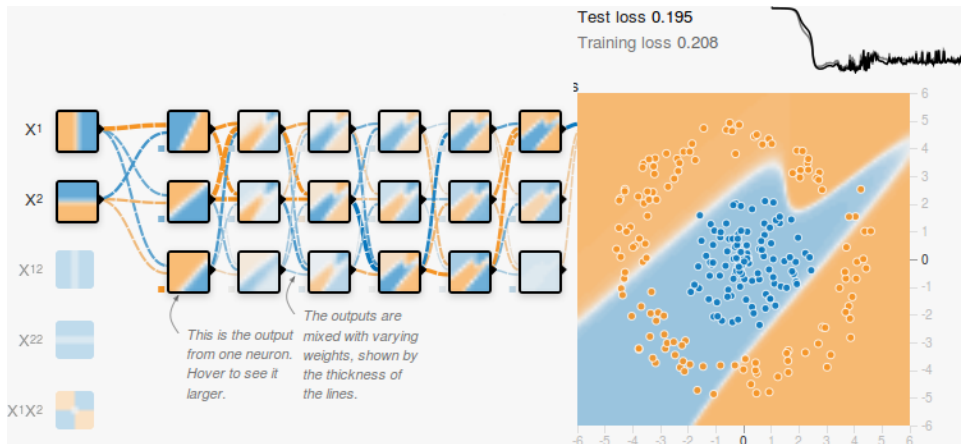
$$1x_1^2 + 1x_2^2 - 1 < 0$$

Bad Features & Low Depth

Test loss 0.510
Training loss 0.488

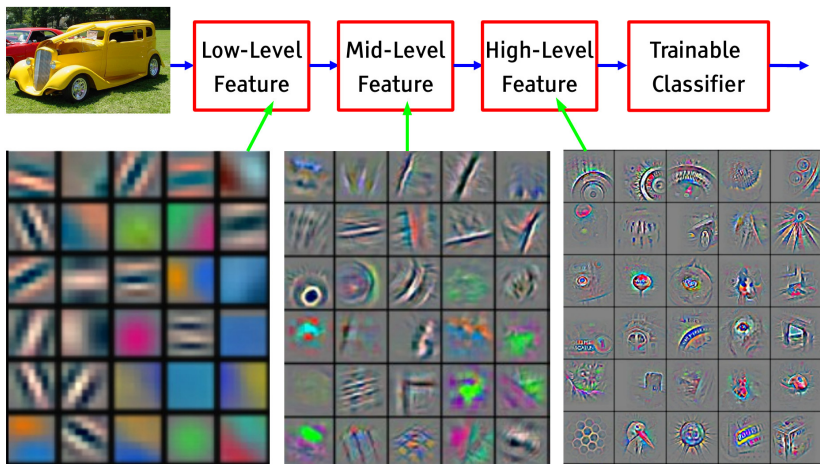


Too Complex NN Fails to Learn



Deep NNs for Image Classification

It's **deep** if it has **more than one stage** of non-linear feature transformation



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

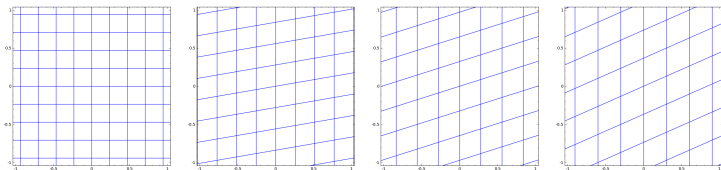
Representation Learning

Representation Learning

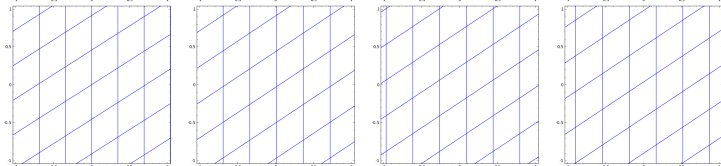
- ▶ We saw DL finding useful features.
- ▶ We can think of these features as *new coordinates*.
- ⇒ NNs are learning how to represent the input to make it linearly separable.

One Layer $\tanh(Wx + b)$, $2D \rightarrow 2D$

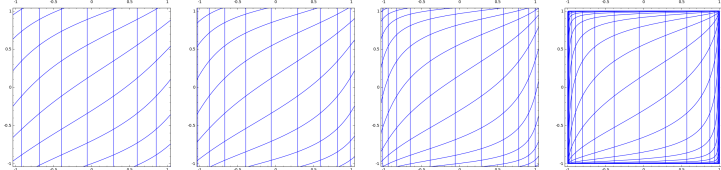
Skew:
 W



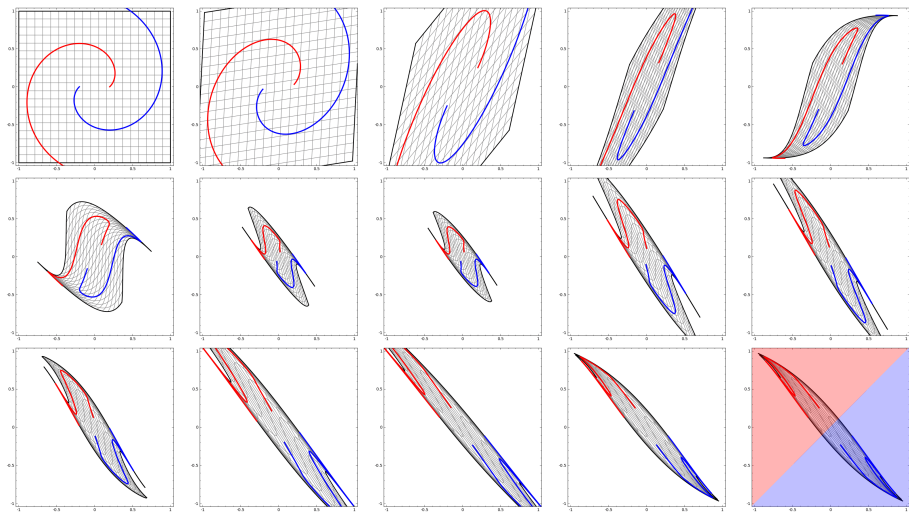
Transpose:
 b



Non-lin.:
 \tanh



Four Layers, Disentangling Spirals



Animation by <http://colah.github.io/posts/2014-03-NN-Manifolds-Topology/>

Back to the Realm of Text

Processing Text with NNs

- Map each word to a vector of 0s and 1s (“1-hot repr.”):

$$\text{cat} \mapsto (0, 0, \dots, 0, 1, 0, \dots, 0)$$

- Sentence is then a matrix:

		the	cat	is	on	the	mat
↑	a	0	0	0	0	0	0
	about	0	0	0	0	0	0

	cat	0	1	0	0	0	0

	is	0	0	1	0	0	0

	on	0	0	0	1	0	0

	the	1	0	0	0	1	0

↓	zebra	0	0	0	0	0	0

Processing Text with NNs

- Map each word to a vector of 0s and 1s (“1-hot repr.”):

$$\text{cat} \mapsto (0, 0, \dots, 0, 1, 0, \dots, 0)$$

- Sentence is then a matrix:

		the	cat	is	on	the	mat
	↑	a	0	0	0	0	0
		about	0	0	0	0	0
	
		cat	0	1	0	0	0
Vocabulary size:	
1.3M English		is	0	0	1	0	0
2.2M Czech	
		on	0	0	0	1	0
	
		the	1	0	0	0	1
	
	↓	zebra	0	0	0	0	0

Processing Text with NNs

- Map each word to a vector of 0s and 1s (“1-hot repr.”):

$$\text{cat} \mapsto (0, 0, \dots, 0, 1, 0, \dots, 0)$$

- Sentence is then a matrix:

		the	cat	is	on	the	mat
	↑	a	0	0	0	0	0
		about	0	0	0	0	0
	
		cat	0	1	0	0	0
Vocabulary size:	
1.3M English		is	0	0	1	0	0
2.2M Czech	
		on	0	0	0	1	0
	
		the	1	0	0	0	1
	
	↓	zebra	0	0	0	0	0

Main drawback: No relations, all words equally close/far.

Sub-Word Units

- ▶ SMT struggled with productive morphology (>1M wordforms).

nejneobhodpodařovatelnějšími, Donaudampfschiffahrtsgesellschaftskapitän

- ▶ NMT can handle only 30–80k dictionaries.

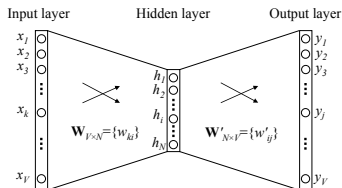
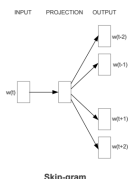
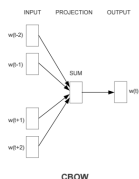
⇒ Resort to sub-word units.

Orig	český politik svezl migranty
Syllables	čes k ý □ po li tik □ sve zl □ mig ran ty
Morphemes	česk ý □ politik □ s vez l □ migrant y
Char Pairs	če sk ý □ po li ti k □ sv ez l □ mi gr an ty
Chars	č e s k ý □ p o l i t i k □ s v e z l □ m i g r a n t y
BPE 30k	český politik s@@ vez@@ l mi@@ granty

BPE (Byte-Pair Encoding) uses n most common substrings (incl. frequent words).

Word Embeddings

- ▶ Map each word to a dense vector.
- ▶ In practice 300–2000 dimensions are used, not 1–2M.
 - ▶ The dimensions have no clear interpretation.
- ▶ Embeddings are trained for each particular task.
 - ▶ NNs: The matrix that maps 1-hot input to the first layer.
- ▶ The famous word2vec (Mikolov et al., 2013):
 - ▶ CBOW: Predict the word from its four neighbours.
 - ▶ Skip-gram: Predict likely neighbours given the word.

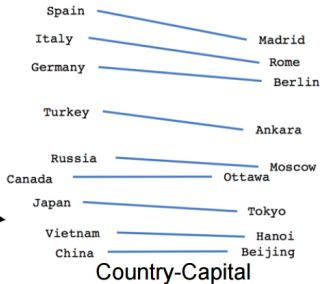
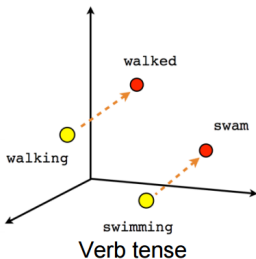
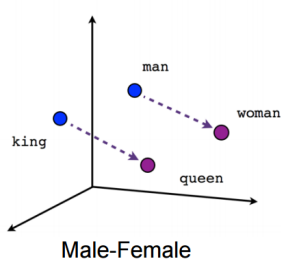


Right: CBOW with just a single-word context
(<http://www-personal.umich.edu/~ronxin/pdf/w2vexp.pdf>)

Continuous Space of Words

Word2vec embeddings show interesting properties:

$$v(\text{king}) - v(\text{man}) + v(\text{woman}) \approx v(\text{queen}) \quad (5)$$



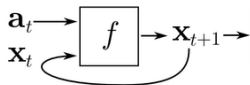
- ▶ Testset by Mikolov et al. (2013) very limited:
 - ▶ “Semantic”: only 3 relation types.
 - ▶ “Syntactic”: only 313 distinct word pairs.
- ▶ Kocmi and Bojar (2016) extend the morphosyntactic part.

Illustrations from <https://www.tensorflow.org/tutorials/word2vec>

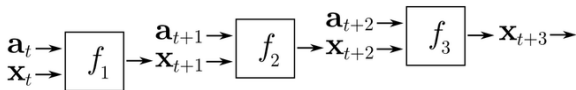
Variable-Length Inputs

Variable-length input can be handled by recurrent NNs:

- ▶ Reading one input symbol at a time.
 - ▶ The same (trained) transformation used repeatedly.
- ▶ Unroll in time (up to a fixed length limit).
- ▶ “Back-propagation through time”.



⇩ unfold through time ⇩



$$\begin{aligned}x_{t+1} &= f(\text{concat}(a_t, x_t)) \\ &= \tanh(W\text{concat}(a_t, x_t) + b)\end{aligned}\tag{6}$$

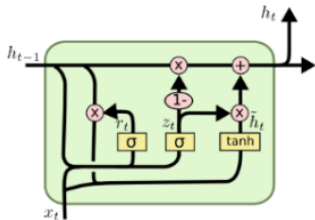
Better RNN Cells

Reusing the same weights easily leads to:

- ▶ exploding gradients (solve by clipping)
- ▶ vanishing gradients

Gating: only part of input/output used/updated:

- ▶ LSTM (Long Short-Term Memory Cells, Hochreiter and Schmidhuber (1997)).
- ▶ GRU (Gated Recurrent Unit Cells, Chung et al. (2014)).



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

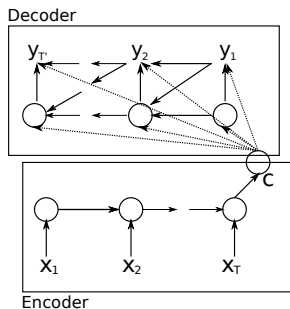
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Important: h_{t-1} and h_t live in the same vector space.

NNs as Translation Model in SMT

Cho et al. (2014) propose:

- ▶ encoder-decoder architecture and
- ▶ GRU unit (name given later by Chung et al. (2014))
- ▶ to score variable-length phrase pairs in PBMT.

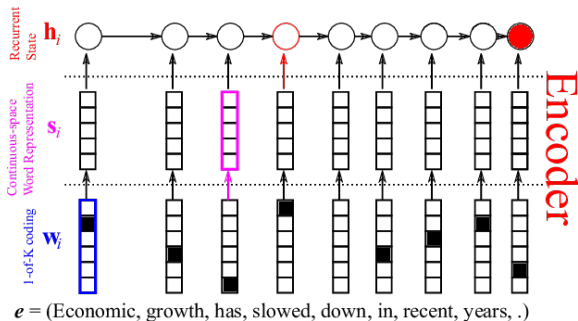


NMT: Sequence to Sequence

Sutskever et al. (2014) use:

- ▶ LSTM RNN encoder-decoder
- ▶ to consume and *produce* variable-length sentences.

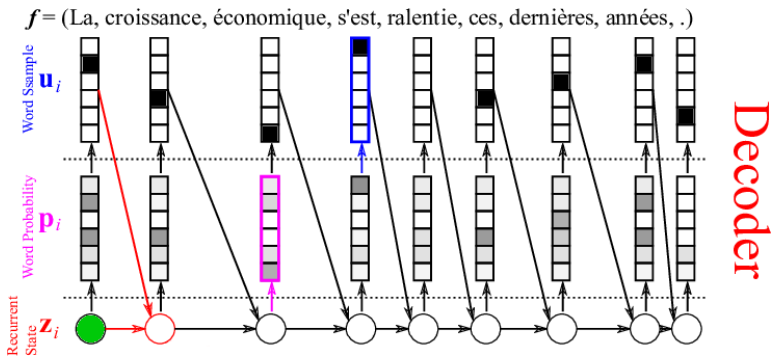
First the Encoder:



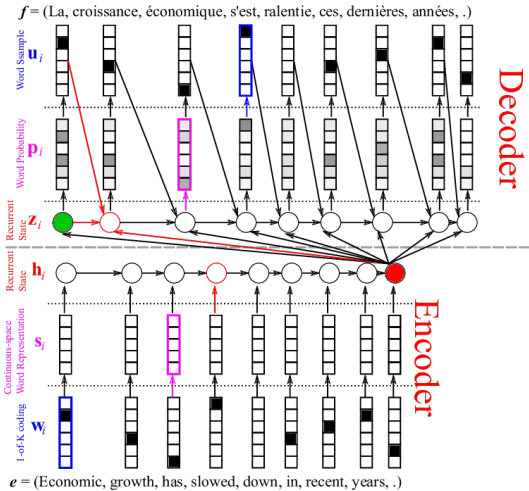
Then the Decoder

Remember: $p(e'_1|f'_1) = p(e_1|f'_1) \cdot p(e_2|e_1, f'_1) \cdot p(e_3|e_2, e_1, f'_1) \dots$

- ▶ Again RNN, producing one word at a time.
- ▶ The produced word fed back into the network.
- ▶ (Word embeddings in the target language used here.)

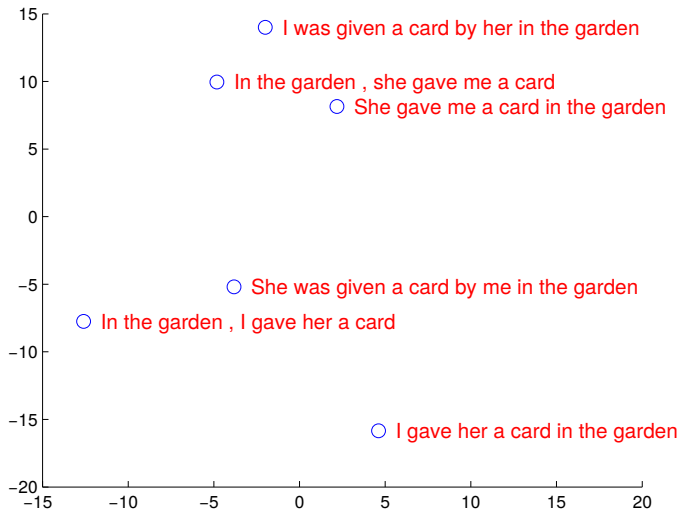


Encoder-Decoder Architecture



<https://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-2/>

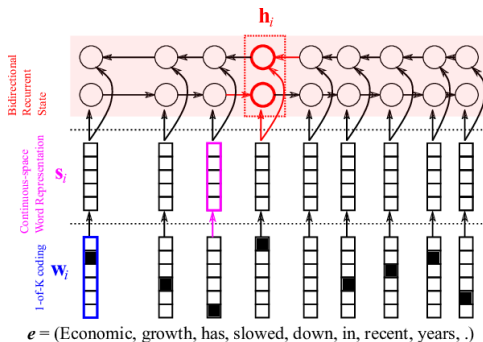
Continuous Space of Sentences



2-D PCA projection of 8000-D space representing sentences (Sutskever et al., 2014).

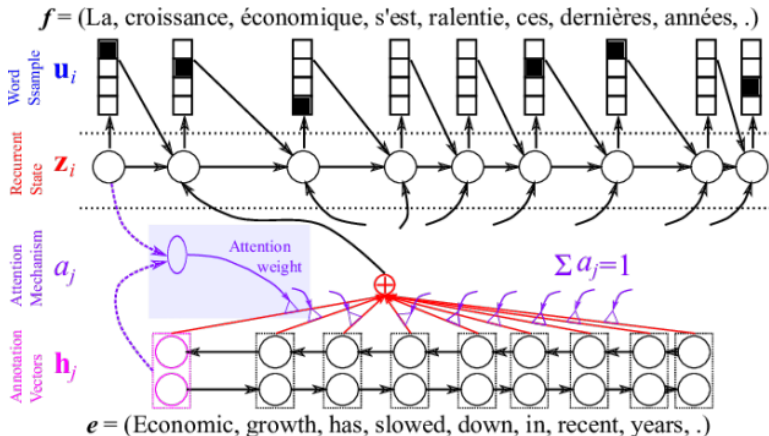
Attention (1/2)

- ▶ Arbitrary-length sentences fit badly into a fixed vector.
 - ▶ Reading input *backward* works better.
 - ... because early words will be more salient.
- ⇒ Use Bi-directional RNN and “attend” to all states h_i .



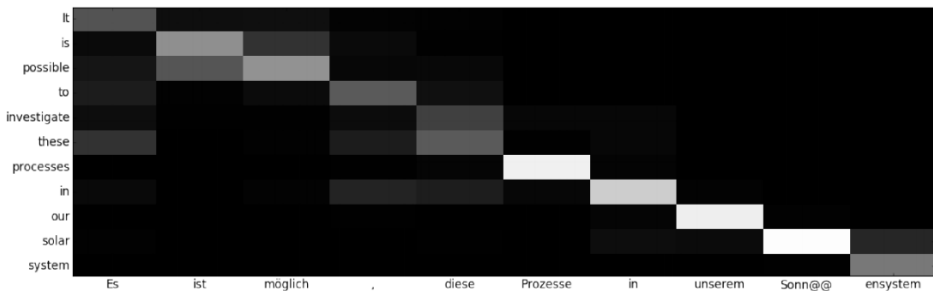
Attention (2/2)

- ▶ Add a sub-network predicting importance of source states at each step.



Attention \approx Alignment

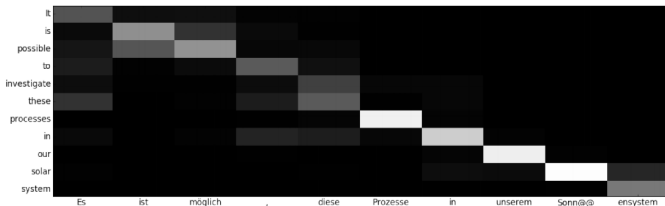
- ▶ We can collect the attention across time.
- ▶ Each column corresponds to one decoder time step.
- ▶ Source tokens correspond to rows.



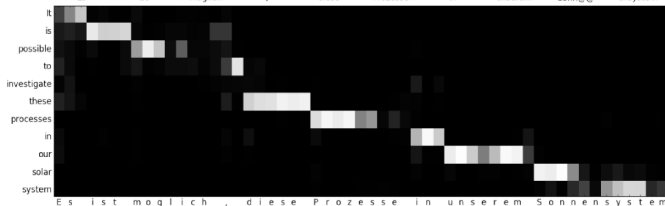
Attending to Characters

“It is possible to investigate these processes in our solar system”

BPE
Decoder

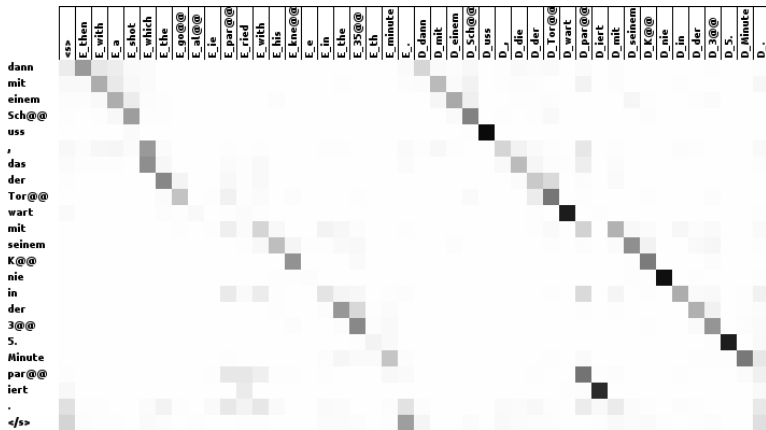


Char
Decoder



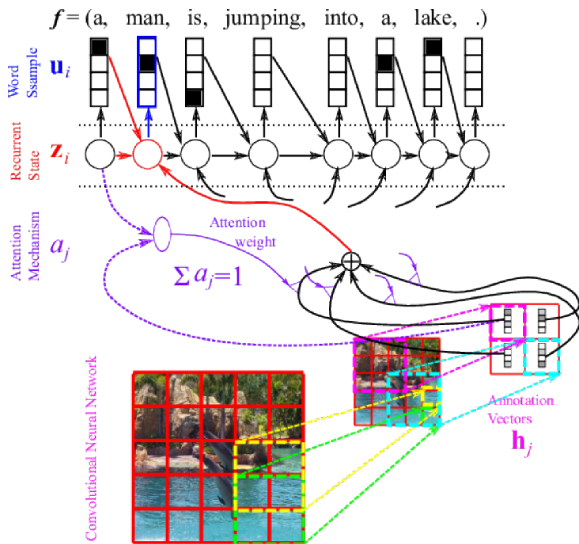
13

Attending to Two Inputs at Once



Niehues et al. (2016) pre-translates with PBMT and then attends both.

Attending to Image Parts



Ultimate Goal of SMT vs. NMT

Goal of “classical” SMT:

Find **minimum translation units** \sim graph partitions:

- ▶ such that they are frequent across many sentence pairs.
- ▶ without imposing (too hard) constraints on reordering.
- ▶ in an unsupervised fashion.

Goal of neural MT:

Avoid minimum translation units.

Come up with a network architecture that:

- ▶ Reads input in as original form as possible.
- ▶ Produces output in as final form as possible.
- ▶ Can be optimized end-to-end *in practice*.

Sample Outputs: NMT or Humans?

SRC 28-Year-Old Chef Found Dead at San Francisco Mall

28letý šéfkuchař Found Dead v San Francisco Mall
Osmadvacetiletý šéfkuchař nalezen mrtev v obchodě
v San Francisku

Sample Outputs: NMT or Humans?

SRC	28-Year-Old Chef Found Dead at San Francisco Mall
MT	28letý šéfkuchař Found Dead v San Francisco Mall
REF	Osmadvacetiletý šéfkuchař nalezen mrtev v obchodě v San Francisku

Sample Outputs: NMT or Humans?

SRC 28-Year-Old Chef Found Dead at San Francisco Mall

MT 28letý šéfkuchař Found Dead v San Francisco Mall

REF Osmadvacetiletý šéfkuchař nalezen mrtev v obchodě v San Francisku

SRC A 28-year-old chef who had recently moved to San Francisco was found dead in the stairwell of a local mall this week.

Osmadvacetiletý kuchař, který se nedávno přestěhoval do San Franciska, byl tento týden nalezen mrtvý na schodišti místního obchodního centra.

Osmadvacetiletý šéfkuchař, který se nedávno přistěhoval do San Franciska, byl tento týden ∅ schodech místního obchodu.

Sample Outputs: NMT or Humans?

SRC 28-Year-Old Chef Found Dead at San Francisco Mall

MT 28letý šéfkuchař Found Dead v San Francisco Mall

REF Osmadvacetiletý šéfkuchař nalezen mrtev v obchodě v San Francisku

SRC A 28-year-old chef who had recently moved to San Francisco was found dead in the stairwell of a local mall this week.

MT Osmadvacetiletý kuchař, který se nedávno přestěhoval do San Franciska, byl tento týden nalezen mrtvý na schodišti místního obchodního centra.

REF Osmadvacetiletý šéfkuchař, který se nedávno přistěhoval do San Franciska, byl tento týden ∅ schodech místního obchodu.

Sample Outputs: NMT or Humans?

SRC A spokesperson for Sons & Daughters **said** they were "shocked and devastated" by his death.

Mluvčí společnosti Sons & Daughters **uvedla**, že jsou jeho smrtí "šokováni a zdrceni".

Mluvčí restaurace Sons & Daughters **řekl**, že jsou jeho smrtí „šokováni a zničení“.

Sample Outputs: NMT or Humans?

SRC A spokesperson for Sons & Daughters **said** they were "shocked and devastated" by his death.

MT Mluvčí společnosti Sons & Daughters **uvedla**, že jsou jeho smrtí "šokováni a zdrceni".

REF Mluvčí restaurace Sons & Daughters **řekl**, že jsou jeho smrtí „šokováni a zničení“.

Sample Outputs: NMT or Humans?

SRC A spokesperson for Sons & Daughters **said** they were "shocked and devastated" by his death.

MT Mluvčí společnosti Sons & Daughters **uvedla**, že jsou jeho smrtí "šokováni a zdrceni".

REF Mluvčí restaurace Sons & Daughters **řekl**, že jsou jeho smrtí „šokováni a zničení“.

SRC "He found an apartment, he was dating a girl," Louis Galicia **a** told KGO.

„Našel si byt, chodil s dívkou,“ řekl Louis Galicia **a** pro KGO.

"Našel si byt, chodil s holkou," řekl Louis Galicia **e** KGO.

Sample Outputs: NMT or Humans?

SRC A spokesperson for Sons & Daughters **said** they were "shocked and devastated" by his death.

MT Mluvčí společnosti Sons & Daughters **uvedla**, že jsou jeho smrtí "šokováni a zdrceni".

REF Mluvčí restaurace Sons & Daughters **řekl**, že jsou jeho smrtí „šokováni a zničení“.

SRC "He found an apartment, he was dating a girl," Louis Galicia **a** told KGO.

REF „Našel si byt, chodil s dívkou,“ řekl Louis Galicia **a** pro KGO.

MT "Našel si byt, chodil s holkou," řekl Louis Galicia **e** KGO.

Luckily ;-), Catastrophic Errors Happen

Same system as above (UEDIN, winner of WMT17):

SRC ... said Frank initially stayed in **hostels**...

MT ... řekl, že Frank původně zůstal v **Budějovicích**...

↳ *Gloss* ... said that Frank initially stayed in **Budweis**...

SRC Most of the **Clintons'** income...

MT Většinu příjmů **Kliniky**...

↳ *Gloss* Most of the income of the **Clinic**...

SRC The 63-year-old has now been made a special repres

MT 63letý **mladík** se nyní stal zvláštním zástupcem...

↳ *Gloss* The 63-year-old **youngster** has now become a speci

Catastrophic Errors Happen (2/2)

SRC Criminal Minds star Thomas Gibson sacked after hitting producer

REF Thomas Gibson, hvězda seriálu Myšlenky zločince, byl propuštěn po té, co uhodil režiséra

MT **Kriminalisté Minsku** hvězdu Thomase Gibsona **vyhostili** po **zásahu** producenta

↳ *Gloss* **Minsk criminal investigators** have **expelled** the star Thomas Gibson after **striking** the producer

SRC ...add to that its long-standing grudge...

REF ...přidejte k tomu svou dlouholetou nenávist...

MT ...přidejte k tomu svou dlouholetou **záštitu**...

↳ *Gloss* ...add to that its long-standing **auspices**...
(grudge = zášť → záštita = auspices)

UEDIN at WMT17

- ▶ Our small annotation of up to 185 sentences.
- ▶ Blind mix: reference or MT.

Real MT was assumed to be:

	OB	DM	DV
MT	142 (76.8 %)	86 (77.5 %)	72 (87.8 %)
didn't know	34 (18.4 %)	9 (8.1 %)	6 (7.3 %)
human	9 (4.9 %)	16 (14.4 %)	4 (4.9 %)
Total	185 (100.0 %)	111 (100.0 %)	82 (100.0 %)

⇒ 10–20% of outputs indistinguishable from humans.

UEDIN at WMT17

- ▶ Our small annotation of up to 185 sentences.
- ▶ Blind mix: reference or MT.

Real MT was assumed to be:

	OB	DM	DV
MT	142 (76.8 %)	86 (77.5 %)	72 (87.8 %)
didn't know	34 (18.4 %)	9 (8.1 %)	6 (7.3 %)
human	9 (4.9 %)	16 (14.4 %)	4 (4.9 %)
Total	185 (100.0 %)	111 (100.0 %)	82 (100.0 %)

⇒ 10–20% of outputs indistinguishable from humans.

	OB	DM	DV
almost flawless	17 (9.19 %)	2 (1.80 %)	0 (0 %)
flawless	82 (44.32 %)	37 (33.33 %)	27 (32.93 %)

⇒ 30–50% of outputs flawless or almost flawless.

How Good NMT is Today? (1/2)

Bilingual manual assessment of translation quality in WMT18:

1/10 blocks, 10 items left in block

WMT18SrcDA #1505:Segment #1487

English → Czech (čeština)

Costs are mounting in the case, with hundreds of pages of affidavits, emails and reports by companies including Deloitte, Pitcher Partners and Charter Keck Cramer filed and top barristers including Allan Myers, QC, and senior solicitors retained by both sides.

— Reference text

V tomto případě rostou Chile má deset členů a koncem srpna by měl společností, včetně společností Deloitte, Pitcher Partners a Charty Keck Cramer, a špičkových obhájců včetně Allana Myerse, QC a vyšších právních zástupců, které si ponechaly obě strany.

— Candidate translation



— How accurately does the above candidate text convey the original semantics of the source text? Slider ranges from Not at all (left) to Perfectly (right).

Reset

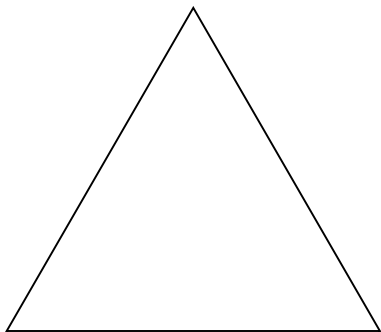
Submit

How Good NMT is Today? (2/2)

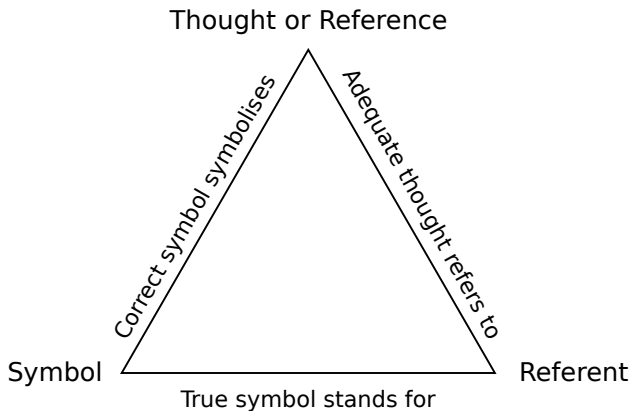
	Ave. %	Ave. z	System
1	84.4	0.667	CUNI-TRANSFORMER
2	79.8	0.521	UEDIN
	78.6	0.483	Reference
4	68.1	0.128	ONLINE-B
5	59.4	-0.178	ONLINE-A
6	54.1	-0.354	ONLINE-G

Caveats:

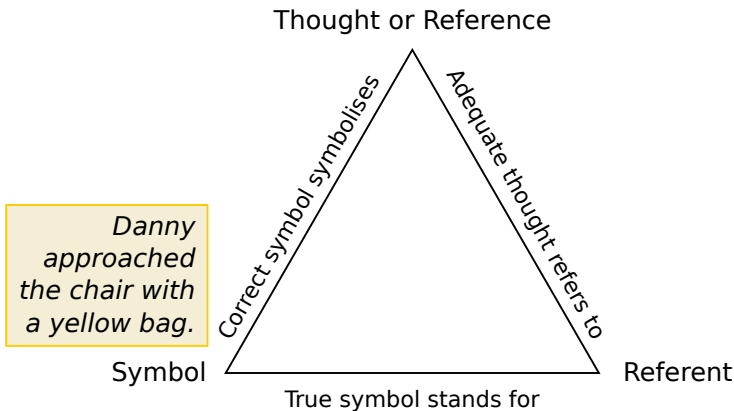
- ▶ Humans translated whole documents, MT individual segments.
- ▶ Evaluation was done for *individual segments*.



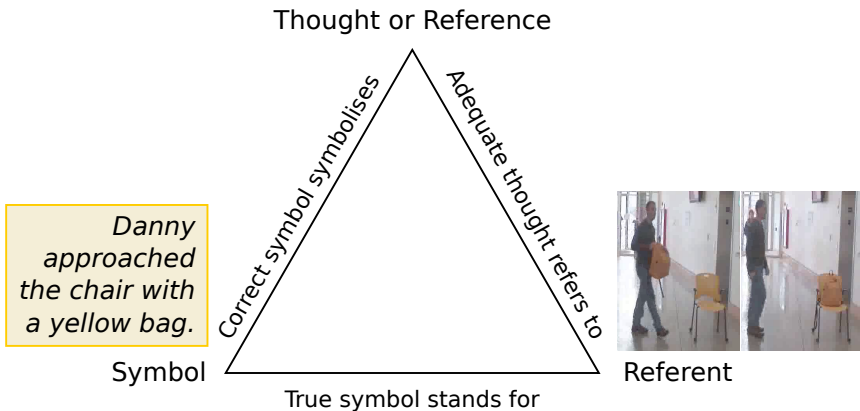
Semiotic Triangle by Ogden and Richards



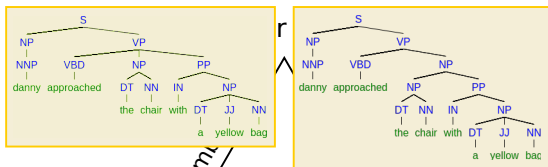
Semiotic Triangle by Ogden and Richards



Semiotic Triangle by Ogden and Richards



Semiotic Triangle by Ogden and Richards



*Danny
approached
the chair with
a yellow bag.*

Symbol

Correct symbol symbol

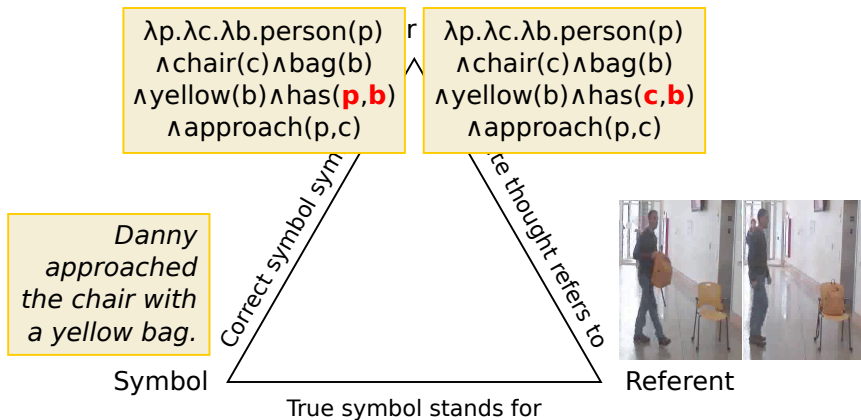
thought refers to

True symbol stands for

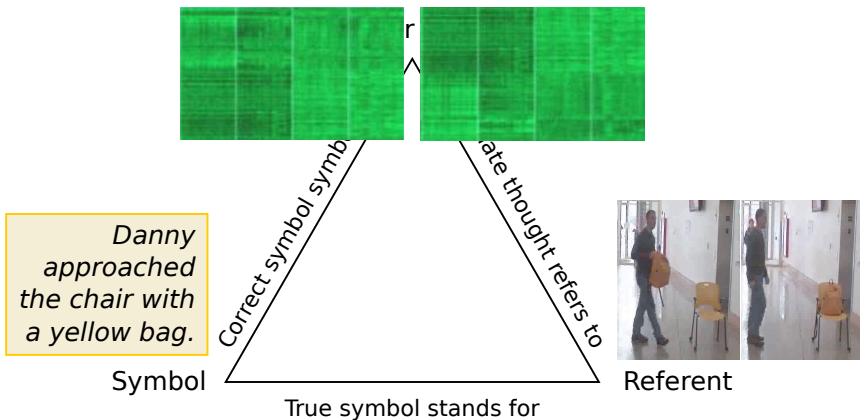


Referent

Semiotic Triangle by Ogden and Richards



Semiotic Triangle by Ogden and Richards



The Proposal for the Next Decade

Train NNs (e.g. NMT) systems for some tasks, and:

A Analyze the meaning **as learnt by the network**, e.g. by:

- ▶ Observing activations, attentions...
- ▶ Lobotomy (Li et al., 2016).
- ▶ Exploring representation space (t-SNE, PCA, ...).
- ▶ Attaching an NN to see if it can infer:
 - ▶ POS or morphology from NMT (Belinkov et al., 2017)
 - ▶ Subject-Verb agreement (Linzen et al. TACL/EACL 2017)
 - ▶ Syntax (Shi et al., 2016)

B Relate the NN behaviour to human behaviour.

- ▶ Record NN behaviour/performance
- ▶ and human behaviour/performance
(including objective measures like eye-tracking, fMRI)
- ... on the same language processing tasks.
- ▶ and relate them.

See DiCarlo 2013 NIPS Tutorial on Vision for a successful application of the idea.

How Continuous is Space of Sentences?

Stages of Space Mapping:

1. Propose directions of exploration.
2. Generate seed pairs of sentences for each of the directions.
3. Collect specimens along the proposed directions:
 - ▶ interpolation, a “sentence in between”,
 - ▶ extrapolation, “a sentence further in the hinted direction”.
 - ▶ Allow people to say “impossible”.
4. Validate the relations.
5. Create the partially ordered set.
6. Search for a manifold covering the ordered set.

Work in progress with Chris Callison-Burch.

Directions of Exploration (1/2)

- ▶ Politeness.
- ▶ Tense.
- ▶ Verity: How much the speaker believes the message.
- ▶ Modality: Willingness/Ability of the speaker to do it.
- ▶ “Counting” / Generic Numerals, Scalar adjectives.
 - ▶ I saw a handful of people there. / a big crowd / a massive crowd.
 - ▶ freezing / cold / chilly
- ▶ “Negation”, but not only reversing the main predicate.
- ▶ Complexity / simplicity, Length.

Directions of Exploration (2/2)

- ▶ Specificity / Generality, Vagueness.
 - ▶ Geese fly / Geese migrate / Geese migrate south / The Canadian geese flew over the pond at friendly Farms in their southward migration.
 - ▶ Hammer the hook into the wall. / Put the hook on the wall. / Do the thingy in there.
- ▶ Contextual boundness.
 - ▶ Give it to him. / Give the parcel to the man at the counter. / Give your parcel to the operator at the post office.
- ▶ High/low style/English/class.
 - ▶ Hey y'all it's a nice day ain't it?
 - ▶ Greetings! Lovely weather we are having.

Thanks to Sarka Zikanova for some of the ideas.

Looking forward for any other ideas you can suggest.

First Results of Getting Pairs

Can you please give me a minute?

Close the door.

Can you help me find something?

May I talk to Mary?

I'm sorry-I don't believe we have met.

Can you move so I can see the screen?

Will you kindly exit?

Would you please get the mail?

Can I help you?

Can you please help me with this?

Can you make me breakfast?

I tried to call were you busy?

Could you leave me alone?

Close the damn door man

I need you to help me get some

Is Mary here?

Who the hell are you?

You aren't made of glass, you k

I do not want you here!

Get the mail!

What do you want?

Get over here and help me!

Why are you not making me br

You never answer your phone.

First Results of Midpointing (1/3)

Can you help me find something?

Find this for me.

Help me find something.

Please help me find something.

Will you help me?

Would you help me look?

Your assistance in finding something is required.

I need you to help me get something.

First Results of Midpointing (2/3)

Can you please give me a minute?

Come back later

Give me a minute.

Hey give me a minute.

I'd like a minute alone.

I need a minute to myself.

I need more time.

One minute.

One moment.

Please wait.

Could you leave me alone?

First Results of Midpointing (3/3)

Can you move so I can see the screen?

Blocking the view, friend.

Can you move a bit?

Can you please move?

Could you move a little bit, you're blocking the screen.

Hey can you move.

I can't see, can you move a little?

Move your blocking the screen

Please move.

You aren't made of glass, you know.

Collect All Variations

When will you be done with your food?

Are you finished with your food?

Are you almost done eating?

Are you finished with your food yet?

Can you hurry eating?

Are you done eating yet?

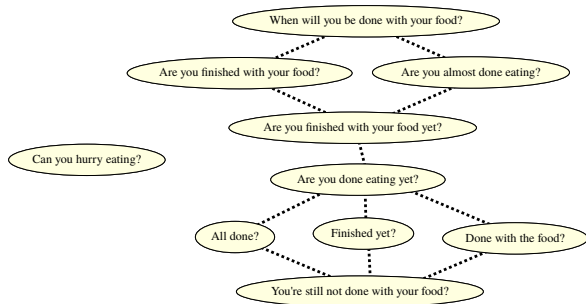
All done?

Finished yet?

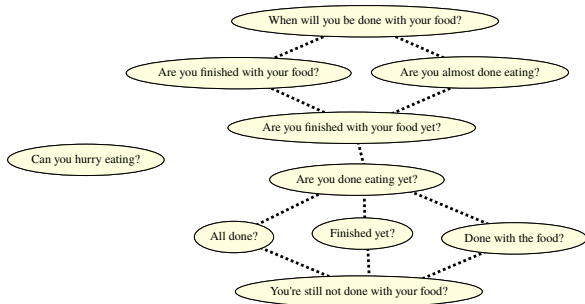
Done with the food?

You're still not done with your food?

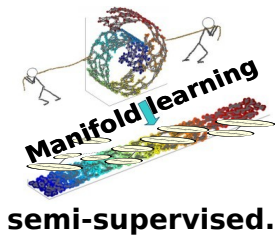
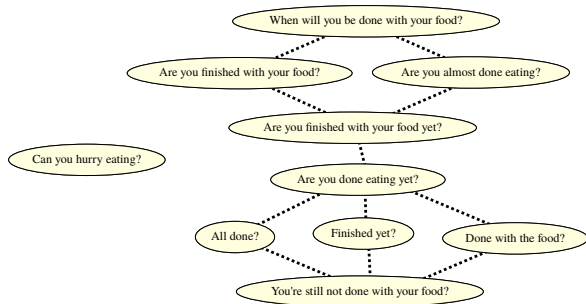
Ask Crowd to Partially Sort Them



Find Methods for Manifold Learning



Match Posets with Learned Manifolds



Counterexamples: Meaning Not Captured

- ▶ Deep learning researchers easily claim that NNs learn the *meaning* of the sentences.
- ▶ This is possible, but not achieved in practice, yet:

English

Czech - detected



English

Czech

Translate

Máma mele maso?
Máma maso mele?
Mele máma maso?
Mele maso máma?
Maso mele máma?
Maso máma mele?



Mum is mincing meat?
Mommy meat?
My mom's meat?
My Flesh Mum?
My mom's meat?
My mom's meat?

Summarization via Machine Translation

Input:

legendární slovenská punkrocková kapela extip se letos vrátila na pódia poté, co vyšla v reedici její debutová deska pekny, škaredý deň, kterou přehraje 1. prosince na sedmičce na strahově. soubor nezanikl, i když bratislavskou punkovou scénu v devadesátých letech rozložily drogy. své zkušenosti s tím má kytarista sveto korbel, který odpovídal na otázky noviněk.

Human Output:

slovenská punková legenda extip se vrátila

Summarization via Machine Translation

Input:

legendární slovenská punkrocková kapela extip se letos vrátila na pódiá poté, co vyšla v reedici její debutová deska pekný, škaredý deň, kterou přehraje 1. prosince na sedmičce na strahově. soubor nezanikl, i když bratislavskou punkovou scénu v devadesátých letech rozložily drogy. své zkušenosti s tím má kytarista sveto korbel, který odpovídal na otázky noviněk.

Human Output:

slovenská punková legenda extip se vrátila

“Sumarized” by Google Transformer Model:

slovenská kapela extip se vrací do prahy

Meaning Understood?

Input:

legendární slovenská punkrocková kapela extip se letos vrátila na pódia poté, co vyšla v reedici její debutová deska pekný, škaredý deň, ktorou prehraje 1. prosince na sedmičke na **strahově**. soubor nezaničl, i když bratislavskou punkovou scénu v devadesátých letech rozložily drogy. své zkušenosti s tím má kytarista sveto korbel, který odpovídal na otázky noviněk.

Human Output:

slovenská punková legenda extip se vrátila

“Sumarized” by Google Transformer Model:

slovenská kapela extip se vrací do **prahy**

Not Understood.

na strahově	slovenská kapela extip se vrací do	prahy
v o2 aréně	slovenská kapela extip se vrací do	prahy
na hradecku	slovenská kapela extip se vrací do	čech
u vajgaru	slovenská kapela extip se vrací do	prahy

Not Understood.

na strahově	slovenská kapela extip se vrací do	prahy
v o2 aréně	slovenská kapela extip se vrací do	prahy
na hradecku	slovenská kapela extip se vrací do	čech
u vajgaru	slovenská kapela extip se vrací do	prahy
ve stromovce	slovenská kapela extip se vrací na scénu. tentokrát kvůli drogám v reedici. s. s. m. m. m. m. m. m. m. m. m. m. m. m. m. m. m. m. i. m. m. . . . m. m. m. m. m. m. m. m. m. m. m. m. m. m. m. m. m. m. m. . m. m. m. m. m. m. m. m. m.	

Shrnutí

Počítačová lingvistika jako zajímavý obor nabízí:

- ▶ Těžké problémy (víceznačnost, vágnost, bohatost jazyka).
- ▶ Nepraktické aplikace (tvaroslovný a větný rozbor).
- ▶ Praktické aplikace (rozpoznání entit, ... strojový překlad).
- ▶ Snadno dostupná data: vstupy a výstupy.
- ▶ Těžce dostupná data: stav mysli mluvčího/posluchače.

Hluboké učení jako nástroj:

- ▶ Skvěle zvládá těžké problémy.
- ▶ Trénováno do přeučení \Rightarrow memoruje, nechápe.

Hluboké učení jako model lidské mysli:

- ▶ “Obecná umělá inteligence” se dá dobře trénovat na jazyce.
- ▶ Mapování lidských a umělých reprezentací.

Náměty na studentské práce (1/2)

Implementační (ročníková / bakalářská práce):

- ▶ Stále častěji je třeba uzavírat *dvojjazyčné smlouvy*.
- ▶ Cílem práce je naimplementovat “dvousloupcový Google Doc s překladem”.

Z překladu (ročníková / bakalářská práce):

- ▶ Ukazuje se, že rusko-arabský překladový model získáme lepší, pokud nezačneme náhodnými čísly, ale anglicko-českým překladem. Čím to?
- ▶ Úkol: natrénovat různé poškozené anglicko-české překlady a zjišťovat, kdy ještě pomáhají tomu rusko-arabskému.

Náměty na studentské práce (2/2)

Z rozpoznávání mluvené řeči (diplomová práce):

- ▶ Tlumočníci i počítač mají velké problémy rozpoznat zkratky a čísla.
- ▶ Často je k projevu dostupný text/prezentace, kde tyto entity jsou.
- ▶ Cílem práce je specializovaný rozpoznávač entit z mluveného slova.
- ▶ Je to “obyčejný” keyword spotting, ale z čerstvých a ne zcela vhodných podkladů.

Z mapování reprezentací (diplomová / disertační práce):

- ▶ Téma načrtnuté výše: mapování variet (manifold learning).

References

- Yonatan Belinkov, Nadir Durrani, Fahim Dalvi, Hassan Sajjad, and James Glass. 2017. What do neural machine translation models learn about morphology? In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 861–872. Association for Computational Linguistics.
- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar, October. Association for Computational Linguistics.
- Junyoung Chung, Çağlar Gülçehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, abs/1412.3555.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November.
- Tom Kocmi and Ondřej Bojar. 2016. SubGram: Extending Skip-gram Word Representation with Substrings. In Petr Sojka, Aleš Horák, Ivan Kopeček, and Karel Pala, editors, *Text, Speech, and Dialogue: 19th International Conference, TSD 2016*, number 9924 in Lecture Notes in Computer Science, pages 182–189, Cham / Heidelberg / New York / Dordrecht / London. Masaryk University, Springer International Publishing.
- Jiwei Li, Will Monroe, and Dan Jurafsky. 2016. Understanding neural networks through representation erasure. *CoRR*, abs/1612.08220.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.
- Jan Niehues, Eunah Cho, Thanh-Le Ha, and Alex Waibel. 2016. Pre-translation for neural machine translation. *CoRR*, abs/1610.05243.
- Xing Shi, Inkit Padhi, and Kevin Knight. 2016. Does string-based neural mt learn source syntax? 