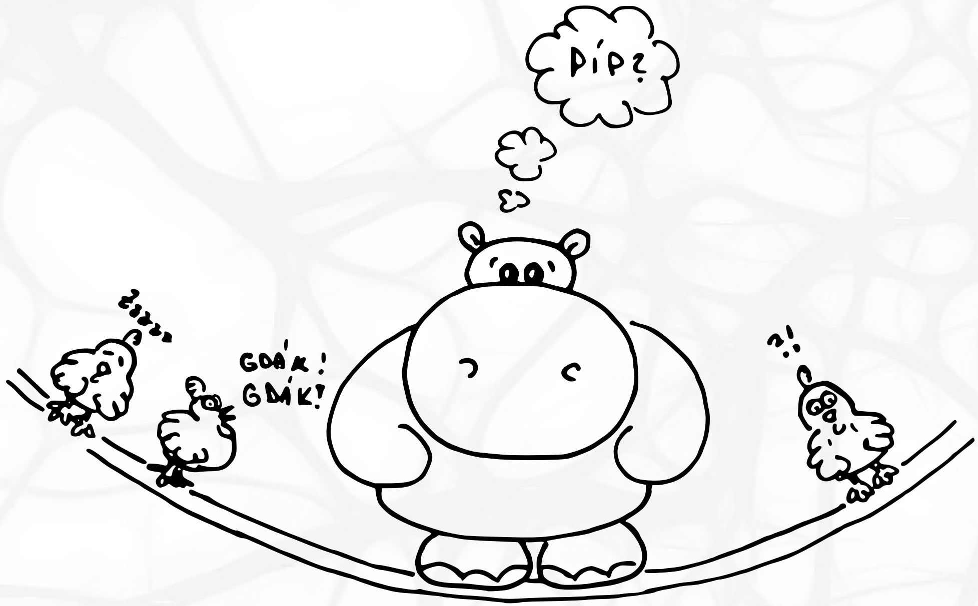
Deep Learning

Deep Learning

Improvements in many areas, matching or exceeding human performance in a lot of them.

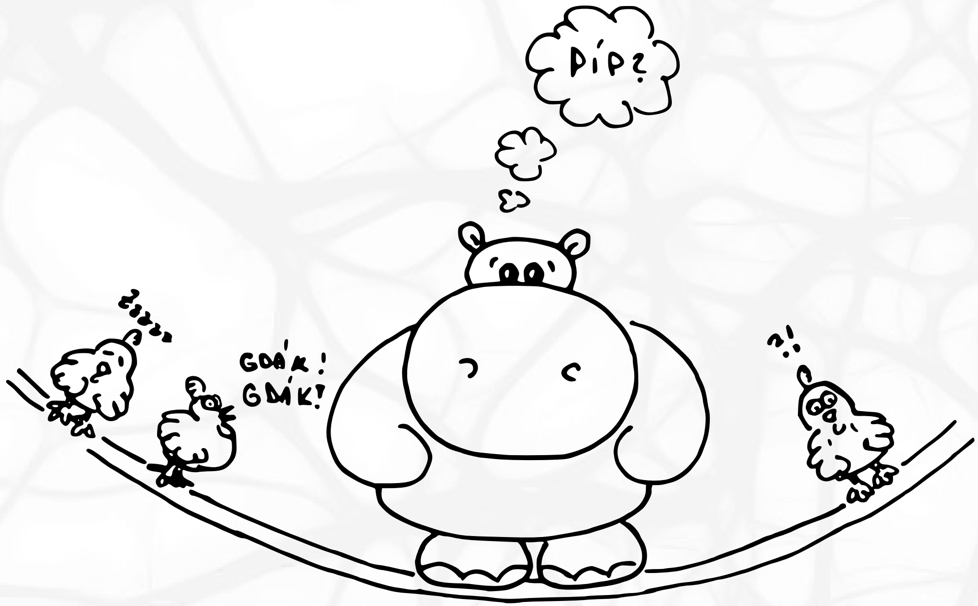
Deep Learning

Improvements in many areas, matching or exceeding human performance in a lot of them.



Deep Learning

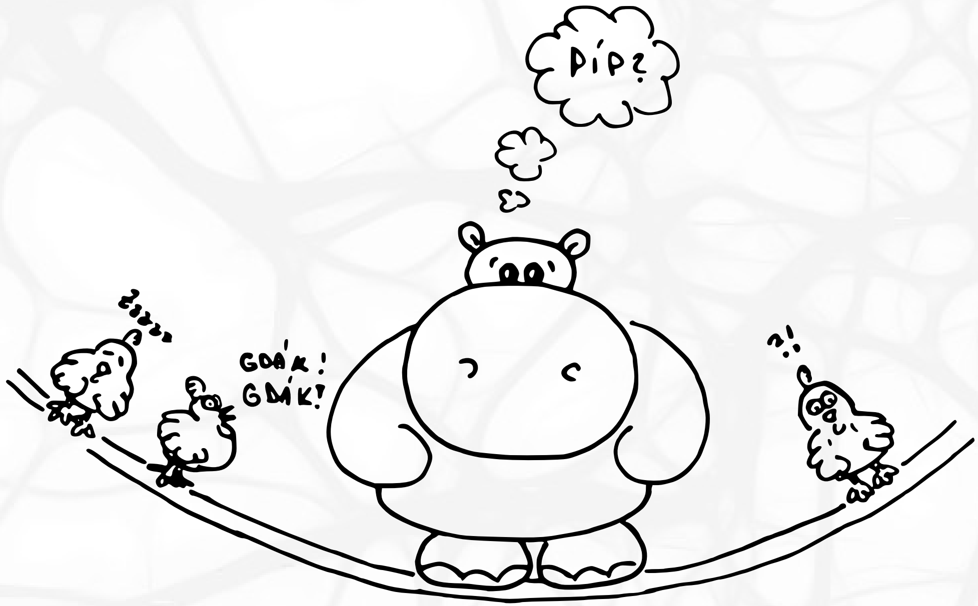
Improvements in many areas, matching or exceeding human performance in a lot of them.



NLP is no exception.

Deep Learning

Improvements in many areas, matching or exceeding human performance in a lot of them.

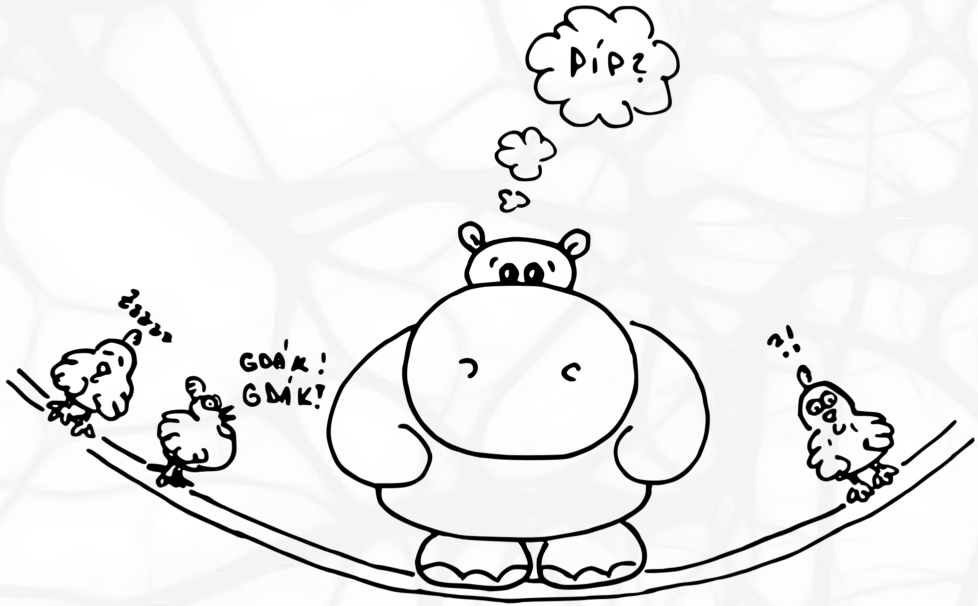


NLP is no exception.

- Improved performance of “basic” NLP tasks, reaching or sometimes even surpassing human performance.

Deep Learning

Improvements in many areas, matching or exceeding human performance in a lot of them.



NLP is no exception.

- Improved performance of “basic” NLP tasks, reaching or sometimes even surpassing human performance.
- Solutions for difficult problems or problems not expected to be solved.

Deep Learning – MorphoDiTa

Method	POS Tags	Lemmas
Original with morphological dictionary	95.55%	97.86%
Deep Neural Network without morphological dictionary		
Deep Neural Network, 8-model ensemble without morphological dictionary		
Deep Neural Network with morphological dictionary		
Deep Neural Network, 8-model ensemble with morphological dictionary		

Deep Learning – MorphoDiTa

Method	POS Tags	Lemmas
Original with morphological dictionary	95.55%	97.86%
Deep Neural Network without morphological dictionary	97.09%	98.37%
Deep Neural Network, 8-model ensemble without morphological dictionary	97.23%	
Deep Neural Network with morphological dictionary	97.37%	
Deep Neural Network, 8-model ensemble with morphological dictionary	97.48%	

Deep Learning – UDPipe

Method	UPOS	XPOS	Feats	Lemmas	UAS	LAS
English – MaltParser guess					*~79.0*	*~75.5*
English – Original	93.50	92.88	94.44	96.10	80.34	77.25
English – Deep NN						
English - CoNLL18 best						
Czech – MaltParser guess					*~85.0*	*~81.5*
Czech - Original	98.23	92.71	91.97	97.82	86.73	83.19
Czech - Deep NN						
Czech - CoNLL18 best						

Deep Learning – UDPipe

Method	UPOS	XPOS	Feats	Lemmas	UAS	LAS
English – MaltParser guess					*~79.0*	*~75.5*
English – Original	93.50	92.88	94.44	96.10	80.34	77.25
English – Deep NN	95.50	95.02	96.06	97.31	85.53	83.02
English - CoNLL18 best	*95.94*	*95.24*	*96.03*	*97.23*	*86.90*	*84.57*
Czech – MaltParser guess					*~85.0*	*~81.5*
Czech - Original	98.23	92.71	91.97	97.82	86.73	83.19
Czech - Deep NN	99.08	97.06	97.01	98.86	92.44	90.42
Czech - CoNLL18 best	*99.07*	*96.95*	*96.89*	*98.71*	*93.44*	*91.68*

Unsupervised Contextualized WEs

Task	Previous SoTA		Baseline	ELMo + Baseline
SQuAD	Liu et al. (2017)	84.4	81.1	85.8
SNLI	Chen et al. (2017)	88.6	88.0	88.7
SRL	He et al. (2017)	81.7	81.4	84.6
Coref	Lee et al. (2017)	67.2	67.2	70.4
NER	Liu et al. (2017)	91.71	90.15	92.22
SST-5	McCann et al. (2017)	53.7	51.4	54.7

Unsupervised Contextualized CLEs

Task	Previous SoTA		Proposed Architecture
NER English	Peters et al. (2018) *Ma and Hovy (2016)*	92.22 *91.21*	93.09
NER German	Lample et al. (2016)	78.76	88.32
Chunking	Peters et al. (2018) *Hashimoto et al. (2016)*	96.37 *95.77*	96.72
POS PTB	Choi (2016)	97.64	97.85

Akbik et al. (COLING 2018); WEs as states of char-level biLSTM LM

Lip Reading

Figure 3. **Top:** Original still images from the BBC lip reading dataset – News, Question Time, Breakfast, Newsnight (from left to right). **Bottom:** The mouth motions for ‘afternoon’ from two different speakers. The network sees the areas inside the red squares.

Figure 3 of "Lip Reading Sentences in the Wild", https://arxiv.org/abs/1611.05358.

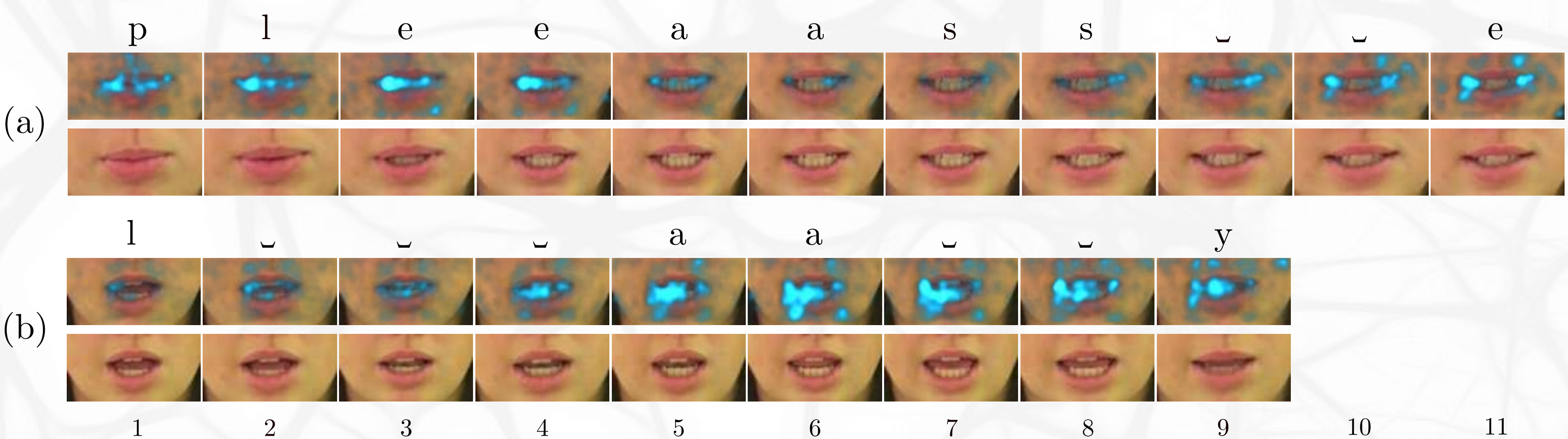


Figure 2 of "LipNet: End-to-end Sentence-level Lipreading", https://arxiv.org/abs/1611.01599.

Lip Reading

Method	SNR	CER	WER	BLEU†
Lips only				
Professional‡	-	58.7%	73.8%	23.8
WAS	-	59.9%	76.5%	35.6
WAS+CL	-	47.1%	61.1%	46.9
WAS+CL+SS	-	42.4%	58.1%	50.0
WAS+CL+SS+BS	-	39.5%	50.2%	54.9
Audio only				
Google Speech API	clean	17.6%	22.6%	78.4
Kaldi SGMM+MMI*	clean	9.7%	16.8%	83.6
LAS+CL+SS+BS	clean	10.4%	17.7%	84.0
LAS+CL+SS+BS	10dB	26.2%	37.6%	66.4
LAS+CL+SS+BS	0dB	50.3%	62.9%	44.6
Audio and lips				
WLAS+CL+SS+BS	clean	7.9%	13.9%	87.4
WLAS+CL+SS+BS	10dB	17.6%	27.6%	75.3
WLAS+CL+SS+BS	0dB	29.8%	42.0%	63.1

Table 5 of "Lip Reading Sentences in the Wild", https://arxiv.org/abs/1611.05358.

Method	Unseen Speakers CER	Unseen Speakers WER	Overlapped Speakers CER	Overlapped Speakers WER
Hearing-Impaired Person (avg)	–	47.7%	–	–
Baseline-LSTM	38.4%	52.8%	15.2%	26.3%
Baseline-2D	16.2%	26.7%	4.3%	11.6%
Baseline-NoLM	6.7%	13.6%	2.0%	5.6%
LipNet	**6.4%**	**11.4%**	**1.9%**	**4.8%**

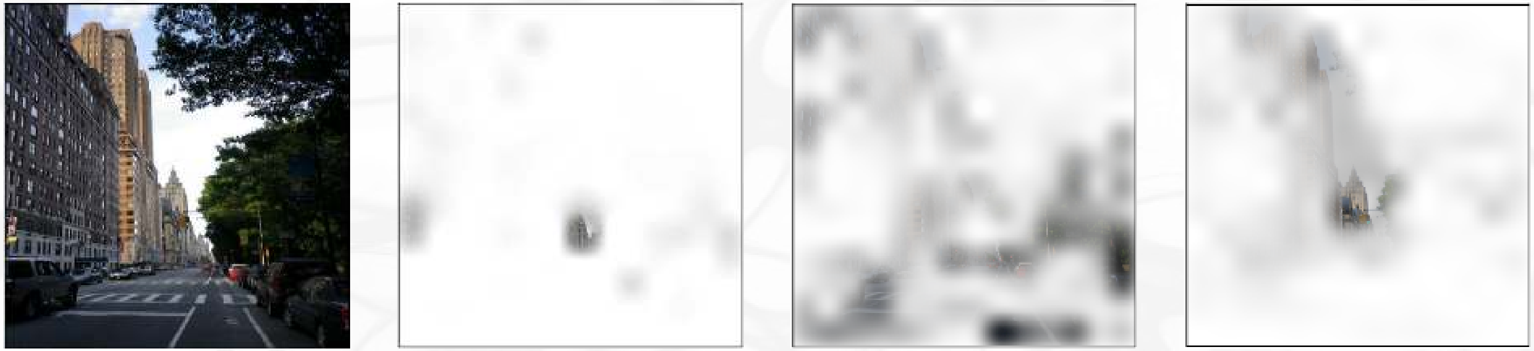
Table 2 of "LipNet: End-to-end Sentence-level Lipreading", https://arxiv.org/abs/1611.01599.

Visual Question Answering

What vegetable is the dog chewing on?
MCB: carrot
GT: carrot

What kind of dog is this?
MCB: husky
GT: husky

What kind of flooring does the room have?
MCB: carpet
GT: carpet



What color is the traffic light?
MCB: green
GT: green

Is this an urban area?
MCB: yes
GT: yes

Where are the buildings?
MCB: in background
GT: on left

Figure 6 of "Multimodal Compact Bilinear Pooling for VQA and Visual Grounding", https://arxiv.org/abs/1606.01847.

Speech Synthesis

System	MOS
Parametric	3.492 ± 0.096
Tacotron (Griffin-Lim)	4.001 ± 0.087
Concatenative	4.166 ± 0.091
WaveNet (Linguistic)	4.341 ± 0.051
Ground truth	4.582 ± 0.053
Tacotron 2 (this paper)	$\mathbf{4.526 \pm 0.066}$

Table 1 of paper "Natural TTS Synthesis by...", https://arxiv.org/abs/1712.05884.

Speech Synthesis

System	MOS
Parametric	3.492 ± 0.096
Tacotron (Griffin-Lim)	4.001 ± 0.087
Concatenative	4.166 ± 0.091
WaveNet (Linguistic)	4.341 ± 0.051
Ground truth	4.582 ± 0.053
Tacotron 2 (this paper)	$\mathbf{4.526 \pm 0.066}$

Table 1 of paper "Natural TTS Synthesis by...", https://arxiv.org/abs/1712.05884.

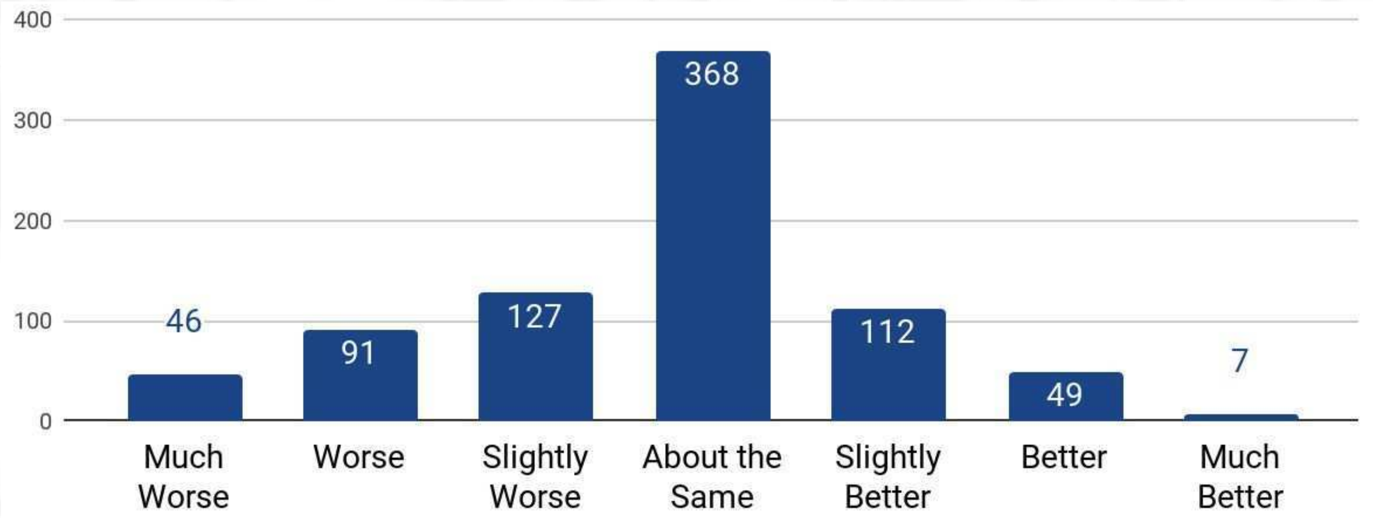
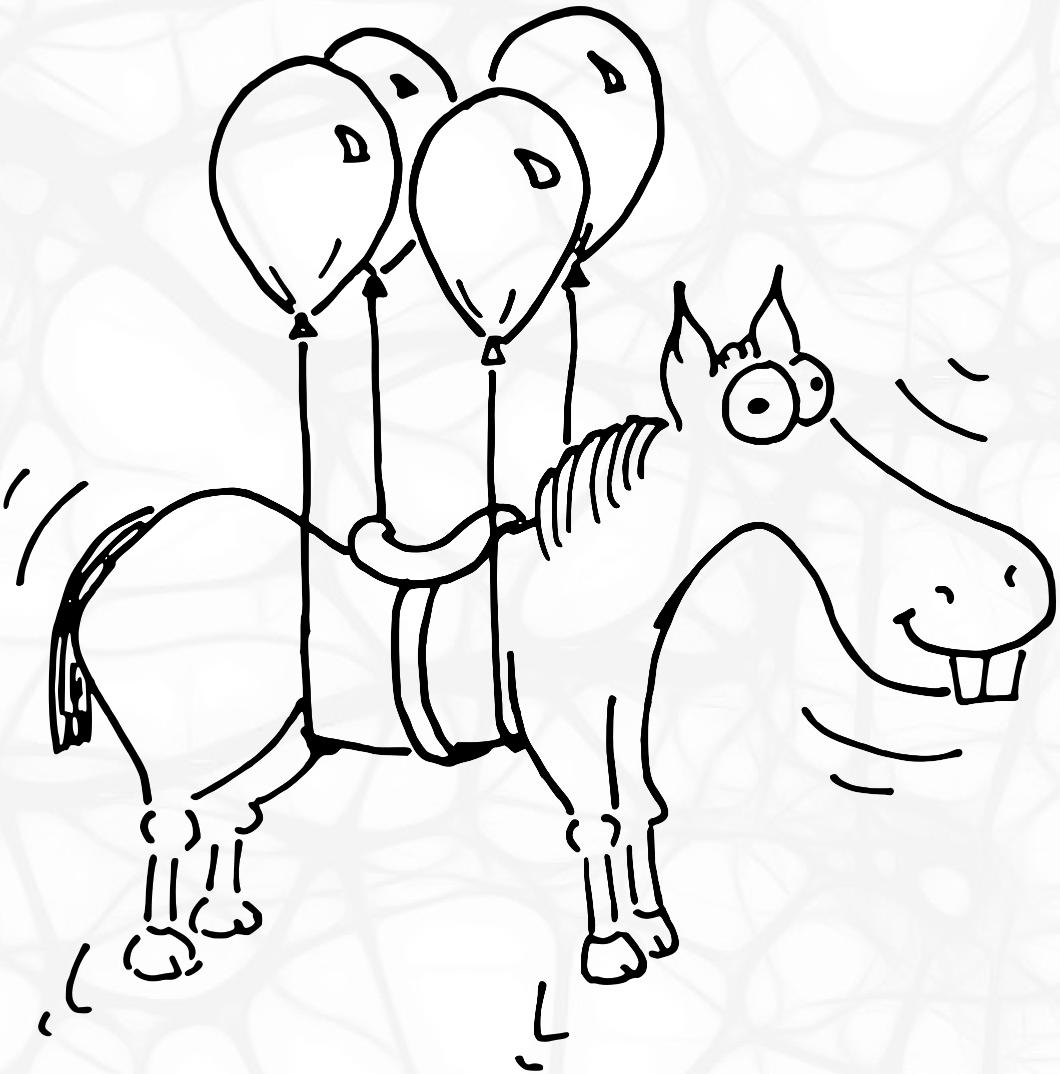


Figure 2 of paper "Natural TTS Synthesis by...", https://arxiv.org/abs/1712.05884.

Unsupervised Word Translation

Unsupervised Word Translation



Unsupervised Word Translation

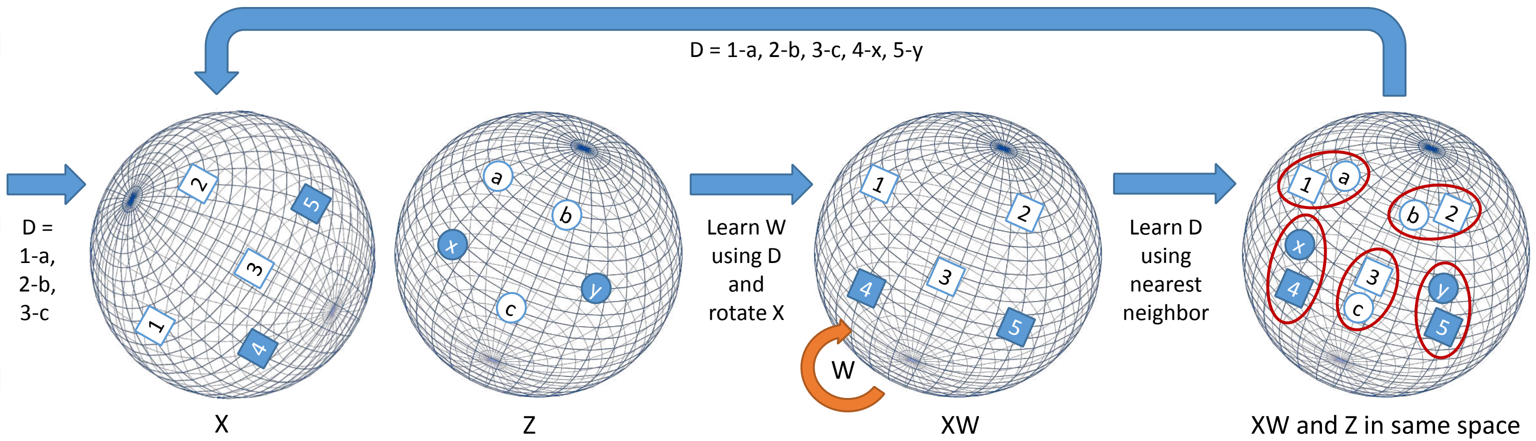


Figure 1 of "Learning bilingual WEs with (almost) no bilingual data", https://aclweb.org/anthology/P17-1042.

Unsupervised Word Translation

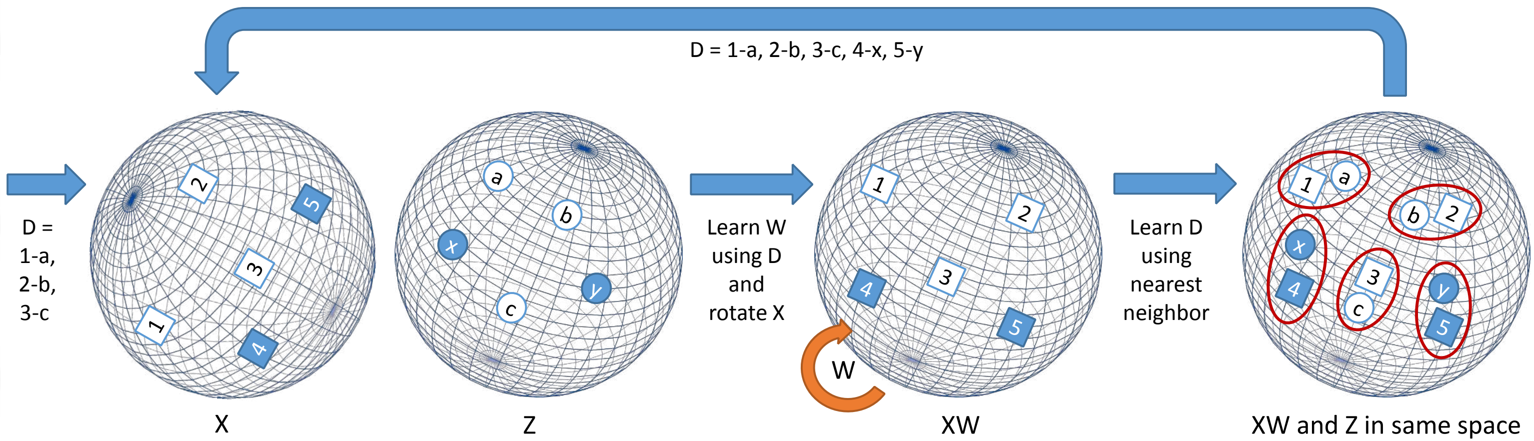


Figure 1 of "Learning bilingual WEs with (almost) no bilingual data", https://aclweb.org/anthology/P17-1042.

	en-es	es-en	en-fr	fr-en	en-de	de-en	en-ru	ru-en	en-zh	zh-en	en-eo	eo-en
Methods with cross-lingual supervision and fastText embeddings												
Procrustes - NN	77.4	77.3	74.9	76.1	68.4	67.7	47.0	58.2	40.6	30.2	22.1	20.4
Procrustes - ISF	81.1	82.6	81.1	81.3	71.1	71.5	49.5	63.8	35.7	**37.5**	29.0	27.9
Procrustes - CSLS	81.4	82.9	81.1	**82.4**	73.5	**72.4**	**51.7**	**63.7**	**42.7**	36.7	**29.3**	25.3
Methods without cross-lingual supervision and fastText embeddings												
Adv - NN	69.8	71.3	70.4	61.9	63.1	59.6	29.1	41.5	18.5	22.3	13.5	12.1
Adv - CSLS	75.7	79.7	77.8	71.2	70.1	66.4	37.2	48.1	23.4	28.3	18.6	16.6
Adv - Refine - NN	79.1	78.1	78.1	78.2	71.3	69.6	37.3	54.3	30.9	21.9	20.7	20.6
Adv - Refine - CSLS	**81.7**	**83.3**	**82.3**	82.1	**74.0**	72.2	44.0	59.1	32.5	31.4	28.2	**25.6**

Table 1: Word translation retrieval P@1 for our released vocabularies in various language pairs. We consider 1,500 source test queries, and 200k target words for each language pair. We use fastText embeddings trained on Wikipedia. NN: nearest n ighbors. ISF: inverted softmax. ('en' is English, 'fr' is French, 'de' is German, 'ru' is Russian, 'zh' is classical Chinese and 'eo' is Esperanto)

Table 1 of "Word Translation Without Parallel Data", https://arxiv.org/abs/1710.04087.

Unsupervised Word Translation

	English to Italian			Italian to English		
	P@1	P@5	P@10	P@1	P@5	P@10
Methods with cross-lingual supervision (WaCky)						
Mikolov et al. (2013b) †	33.8	48.3	53.9	24.9	41.0	47.4
Dinu et al. (2015)†	38.5	56.4	63.9	24.6	45.4	54.1
CCA†	36.1	52.7	58.1	31.0	49.9	57.0
Artetxe et al. (2017)	39.7	54.7	60.5	33.8	52.4	59.1
Smith et al. (2017)†	43.1	60.7	66.4	38.0	58.5	63.6
Procrustes - CSLS	44.9	61.8	66.6	38.5	57.2	63.0
Methods without cross-lingual supervision (WaCky)						
Adv - Refine - CSLS	45.1	60.7	65.1	38.3	57.8	62.8
Methods with cross-lingual supervision (Wiki)						
Procrustes - CSLS	63.7	78.6	81.1	56.3	76.2	80.6
Methods without cross-lingual supervision (Wiki)						
Adv - Refine - CSLS	**66.2**	**80.4**	**83.4**	**58.7**	**76.5**	**80.9**

Table 2: English-Italian word translation average precisions (@1, @5, @10) from 1.5k source word queries using 200k target words. Results marked with the symbol † are from Smith et al. (2017). Wiki means the embeddings were trained on Wikipedia using fastText. Note that the method used by Artetxe et al. (2017) does not use the same supervision as other supervised methods, as they only use numbers in their initial parallel dictionary.

Table 2 of "Word Translation Without Parallel Data", https://arxiv.org/abs/1710.04087.

Unsupervised Machine Translation

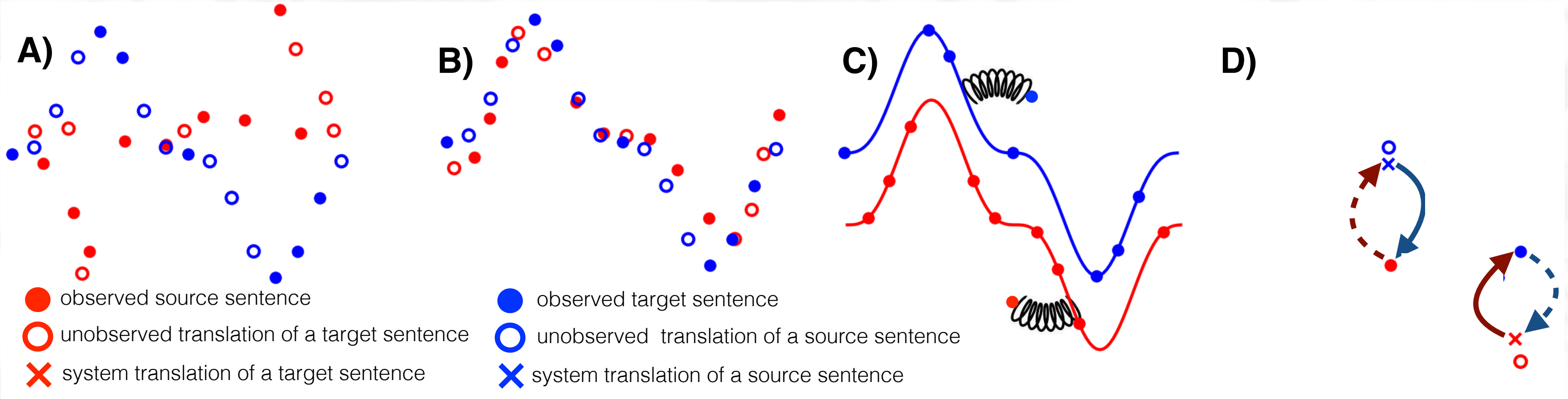


Figure 1 of "Phrase-Based & Neural Unsupervised MT", https://arxiv.org/abs/1804.07755.

Unsupervised Machine Translation

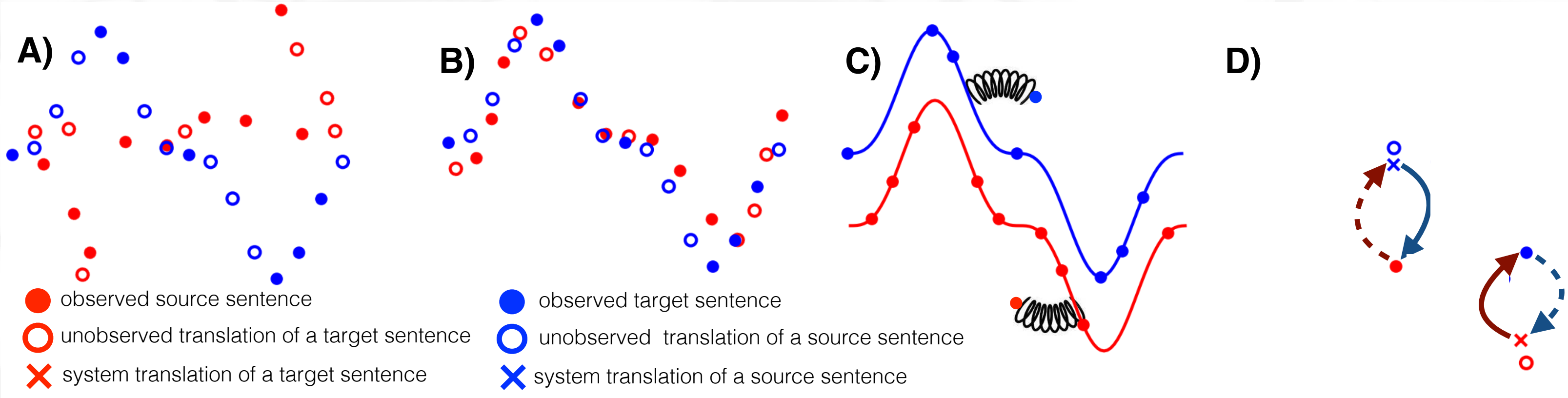


Figure 1 of "Phrase-Based & Neural Unsupervised MT", https://arxiv.org/abs/1804.07755.

- Shared Word Embeddings

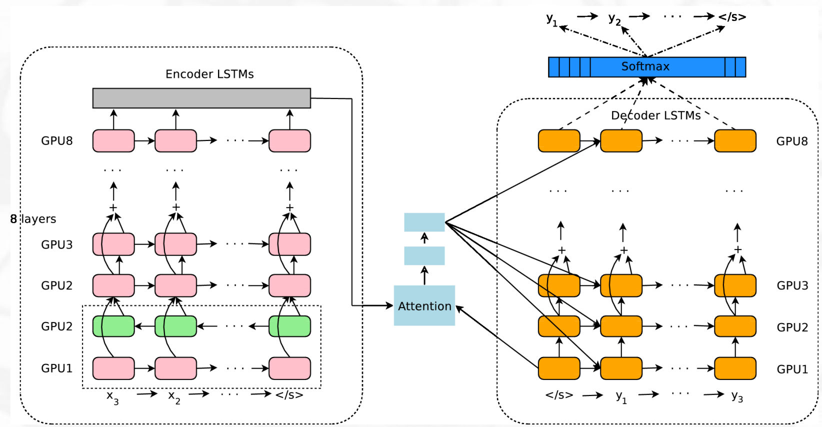


Figure 1 of paper "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation", https://arxiv.org/abs/1609.08144.

Unsupervised Machine Translation

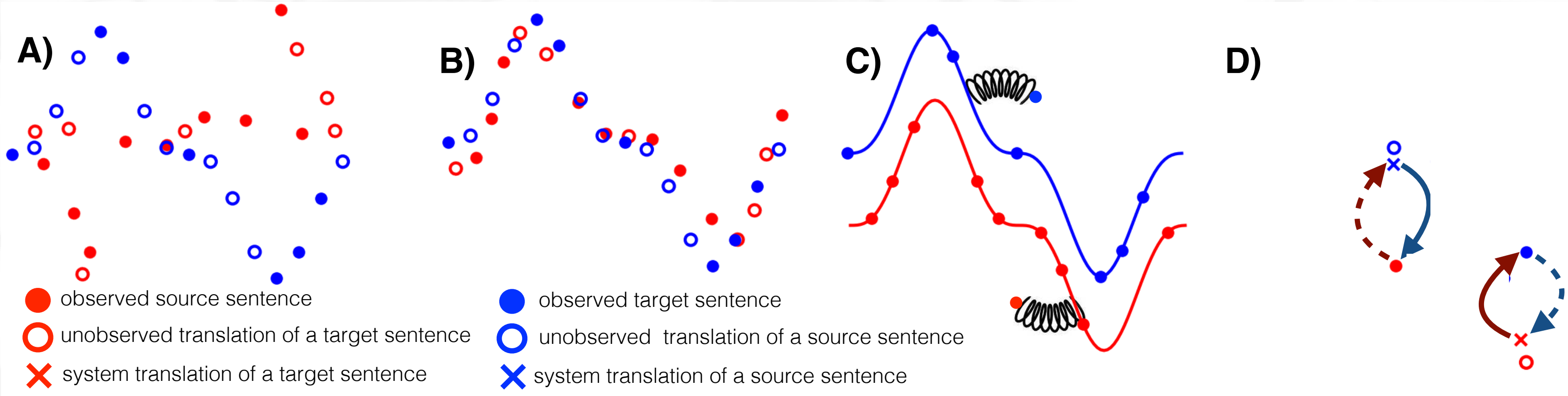


Figure 1 of "Phrase-Based & Neural Unsupervised MT", https://arxiv.org/abs/1804.07755.

- Shared Word Embeddings
- Denoising

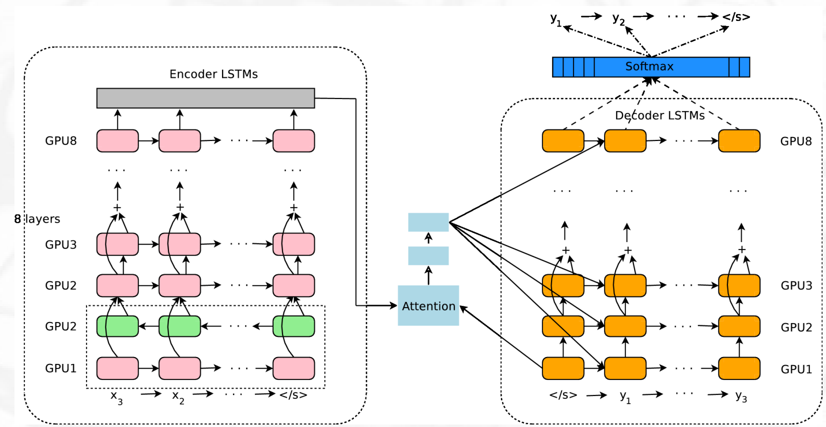


Figure 1 of paper "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation", https://arxiv.org/abs/1609.08144.

Unsupervised Machine Translation

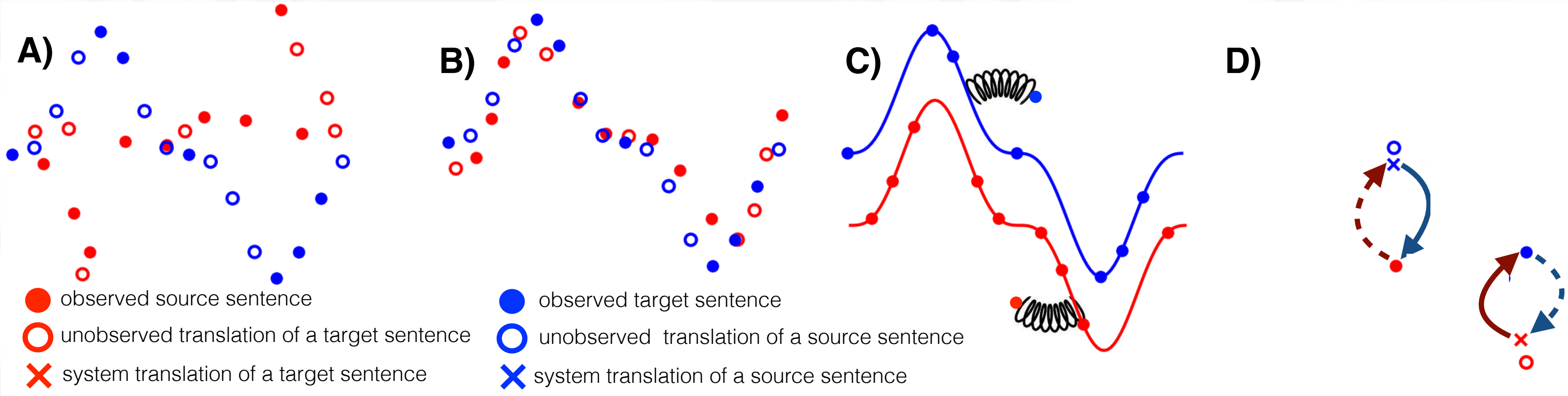


Figure 1 of "Phrase-Based & Neural Unsupervised MT", https://arxiv.org/abs/1804.07755.

- Shared Word Embeddings
- Denoising
- Backtranslation

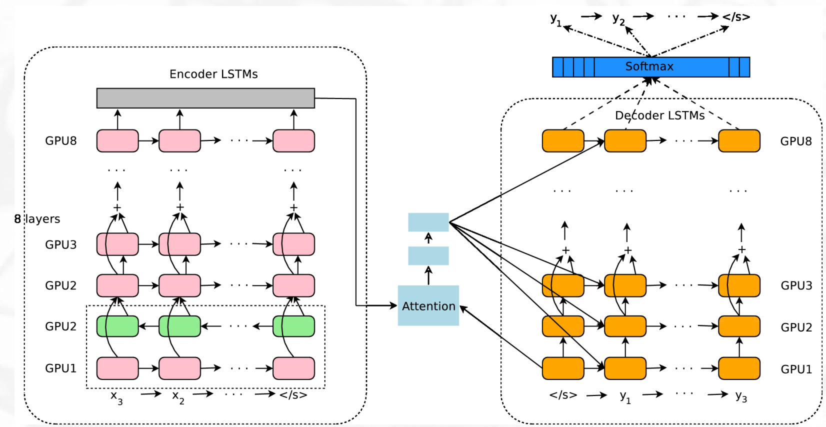


Figure 1 of paper "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation", https://arxiv.org/abs/1609.08144.

Unsupervised Machine Translation

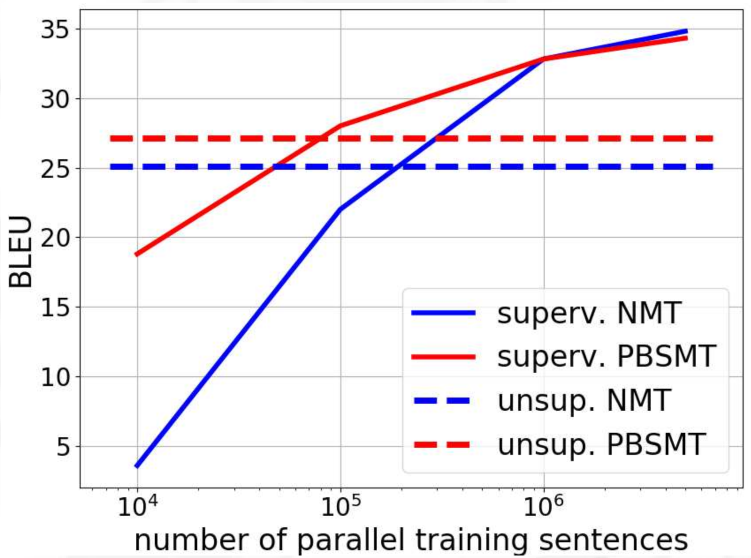


Figure 2: Comparison between supervised and unsupervised approaches on WMT'14 En-Fr, as we vary the number of parallel sentences for the supervised methods.

Figure 2 of "Phrase-Based & Neural Unsupervised MT", https://arxiv.org/abs/1804.07755.

Model	en-fr	fr-en	en-de	de-en
(Artetxe et al., 2018)	15.1	15.6	-	-
(Lample et al., 2018)	15.0	14.3	9.6	13.3
(Yang et al., 2018)	17.0	15.6	10.9	14.6
NMT (LSTM)	24.5	23.7	14.7	19.6
NMT (Transformer)	25.1	24.2	17.2	21.0
PBSMT (Iter. 0)	16.2	17.5	11.0	15.6
PBSMT (Iter. n)	**28.1**	27.2	17.9	22.9
NMT + PBSMT	27.1	26.3	17.5	22.1
PBSMT + NMT	27.6	**27.7**	**20.2**	**25.2**

Table 2: **Comparison with previous approaches.** BLEU score for different models on the $en-fr$ and $en-de$ language pairs. Just using the unsupervised phrase table, and without back-translation (PBSMT (Iter. 0)), the PBSMT outperforms previous approaches. Combining PBSMT with NMT gives the best results.

Table 2 of "Phrase-Based & Neural Unsupervised MT", https://arxiv.org/abs/1804.07755.

Unsupervised Machine Translation

Source	**Je rêve constamment d'eux, peut-être pas toutes les nuits mais plusieurs fois par semaine c'est certain.**
NMT Epoch 1	I constantly dream, but not all nights but by several times it is certain.
NMT Epoch 3	I continually dream them, perhaps not all but several times per week is certain.
NMT Epoch 45	I constantly dream of them, perhaps not all nights but several times a week it 's certain.
PBSMT Iter. 0	I dream of, but they constantly have all those nights but several times a week is too much. "
PBSMT Iter. 2	I had dreams constantly of them, probably not all nights but several times a week it is large.
PBSMT Iter. 8	I dream constantly of them, probably not all nights but several times a week it is certain.
Reference	**I constantly dream of them, perhaps not every night, but several times a week for sure.**

Source	**La protéine que nous utilisons dans la glace réagit avec la langue à pH neutre.**
NMT Epoch 1	The protein that we use in the ice with the language to pH.
NMT Epoch 8	The protein we use into the ice responds with language to pH neutral.
NMT Epoch 45	The protein we use in ice responds with the language from pH to neutral.
PBSMT Iter. 0	The protein that used in the ice responds with the language and pH neutral.
PBSMT Iter. 2	The protein that we use in the ice responds with the language to pH neutral.
PBSMT Iter. 8	The protein that we use in the ice reacts with the language to a neutral pH.
Reference	**The protein we are using in the ice cream reacts with your tongue at neutral pH.**

Source	**Selon Google, les déguisements les plus recherchés sont les zombies, Batman, les pirates et les sorcières.**
NMT Epoch 1	According to Google, there are more than zombies, Batman, and the pirates.
NMT Epoch 8	Google's most wanted outfits are the zombies, Batman, the pirates and the evil.
NMT Epoch 45	Google said the most wanted outfits are the zombies, Batman, the pirates and the witch.
PBSMT Iter. 0	According to Google, fancy dress and most wanted fugitives are the bad guys, Wolverine, the pirates and their minions.
PBSMT Iter. 2	According to Google, the outfits are the most wanted fugitives are zombies, Batman, pirates and witches.
PBSMT Iter. 8	According to Google, the outfits, the most wanted list are zombies, Batman, pirates and witches.
Reference	**According to Google, the highest searched costumes are zombies, Batman, pirates and witches.**

Table 4 of "Phrase-Based & Neural Unsupervised MT", https://arxiv.org/abs/1804.07755.

Many More Successes Outside NLP

- Artificial intelligence – game playing, controlling data centers cooling, …

Many More Successes Outside NLP

- Artificial intelligence – game playing, controlling data centers cooling, …
- Health care – algorithms matching or surpassing humans in diagnostics (95% vs 91% for diabetic retinopathy, 89% vs 73% in tumor localization scores, comparable in analyzing orthopedic trauma radiographs, …)

Many More Successes Outside NLP

- Artificial intelligence – game playing, controlling data centers cooling, …
- Health care – algorithms matching or surpassing humans in diagnostics (95% vs 91% for diabetic retinopathy, 89% vs 73% in tumor localization scores, comparable in analyzing orthopedic trauma radiographs, …)
- …

Many More Successes Outside NLP

- Artificial intelligence – game playing, controlling data centers cooling, …
- Health care – algorithms matching or surpassing humans in diagnostics (95% vs 91% for diabetic retinopathy, 89% vs 73% in tumor localization scores, comparable in analyzing orthopedic trauma radiographs, …)
- …

Short-term Challenges for Us

- start releasing tools using deep NN models

Short-term Challenges for Us

- start releasing tools using deep NN models
  - many technical challenges
    - support GPUs and other emerging HW, handle various RAM sizes, efficient usage of batching, what library/backend to use, backend compilation/distribution, different CUDA versions, support GPUs in REST services, …

Short-term Challenges for Us

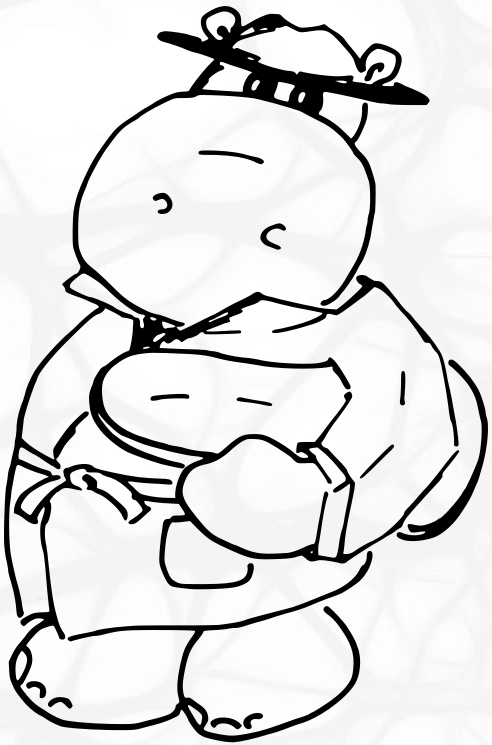
- start releasing tools using deep NN models
  - many technical challenges
    - support GPUs and other emerging HW, handle various RAM sizes, efficient usage of batching, what library/backend to use, backend compilation/distribution, different CUDA versions, support GPUs in REST services, …

- apply deep NN models (even for “solved” tasks) if performance is crucial

Short-term Challenges for Us

- start releasing tools using deep NN models
  - many technical challenges
    - support GPUs and other emerging HW, handle various RAM sizes, efficient usage of batching, what library/backend to use, backend compilation/distribution, different CUDA versions, support GPUs in REST services, ...

- apply deep NN models (even for “solved” tasks) if performance is crucial
- solve tasks that have been “too difficult” so far



Questions