Text *Style* Transfer

Hung-yi Lee 李宏毅

Image Style

Audio Style

male

female

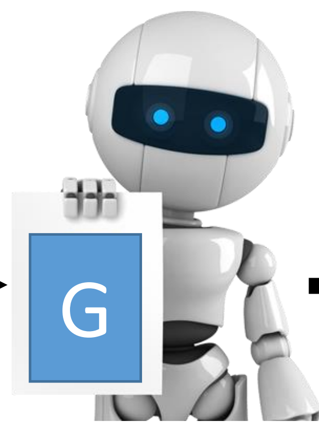
positive sentences

negative sentences

Text Style Transfer

Text Style Transfer

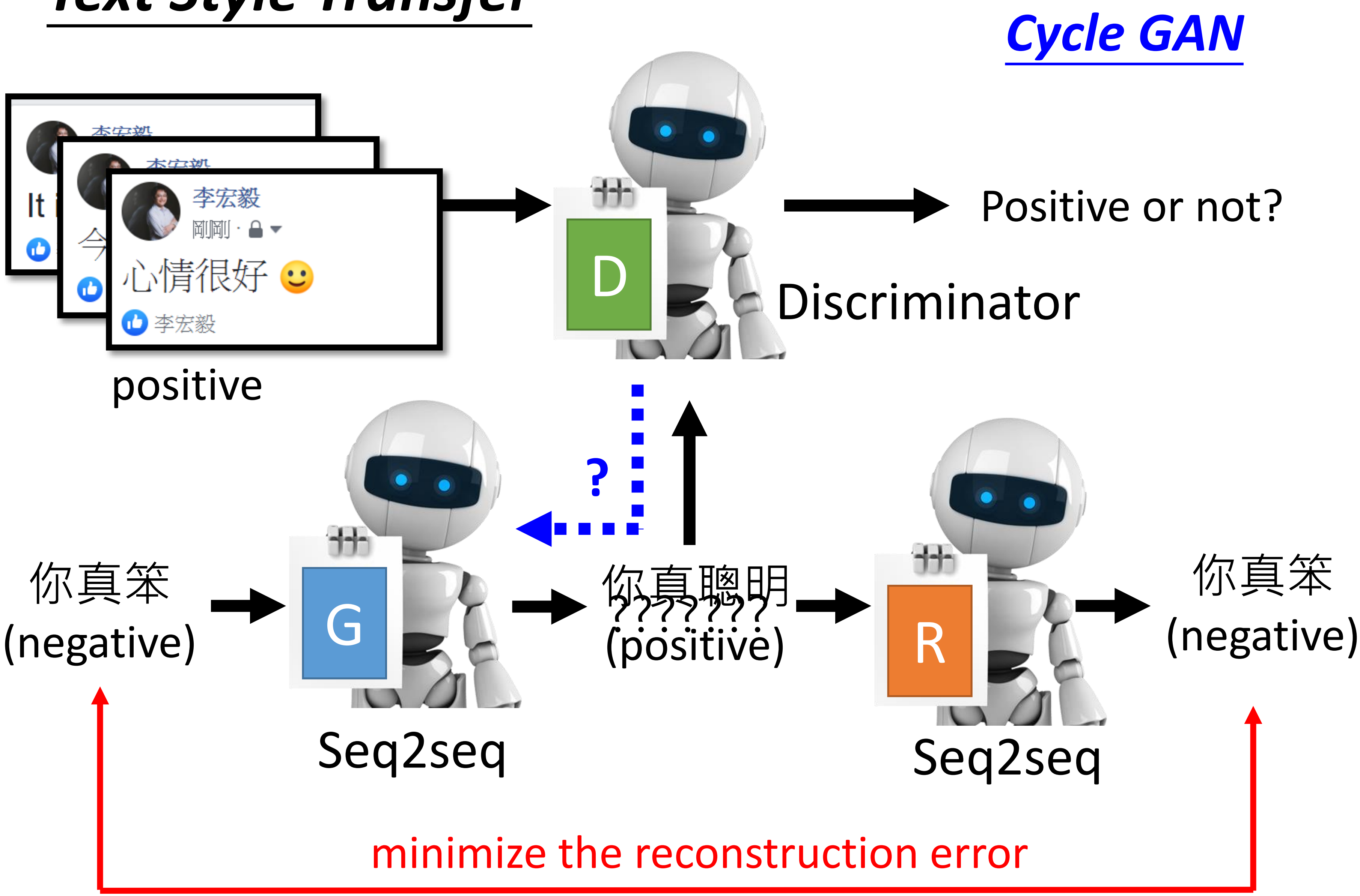
你真笨
(negative)



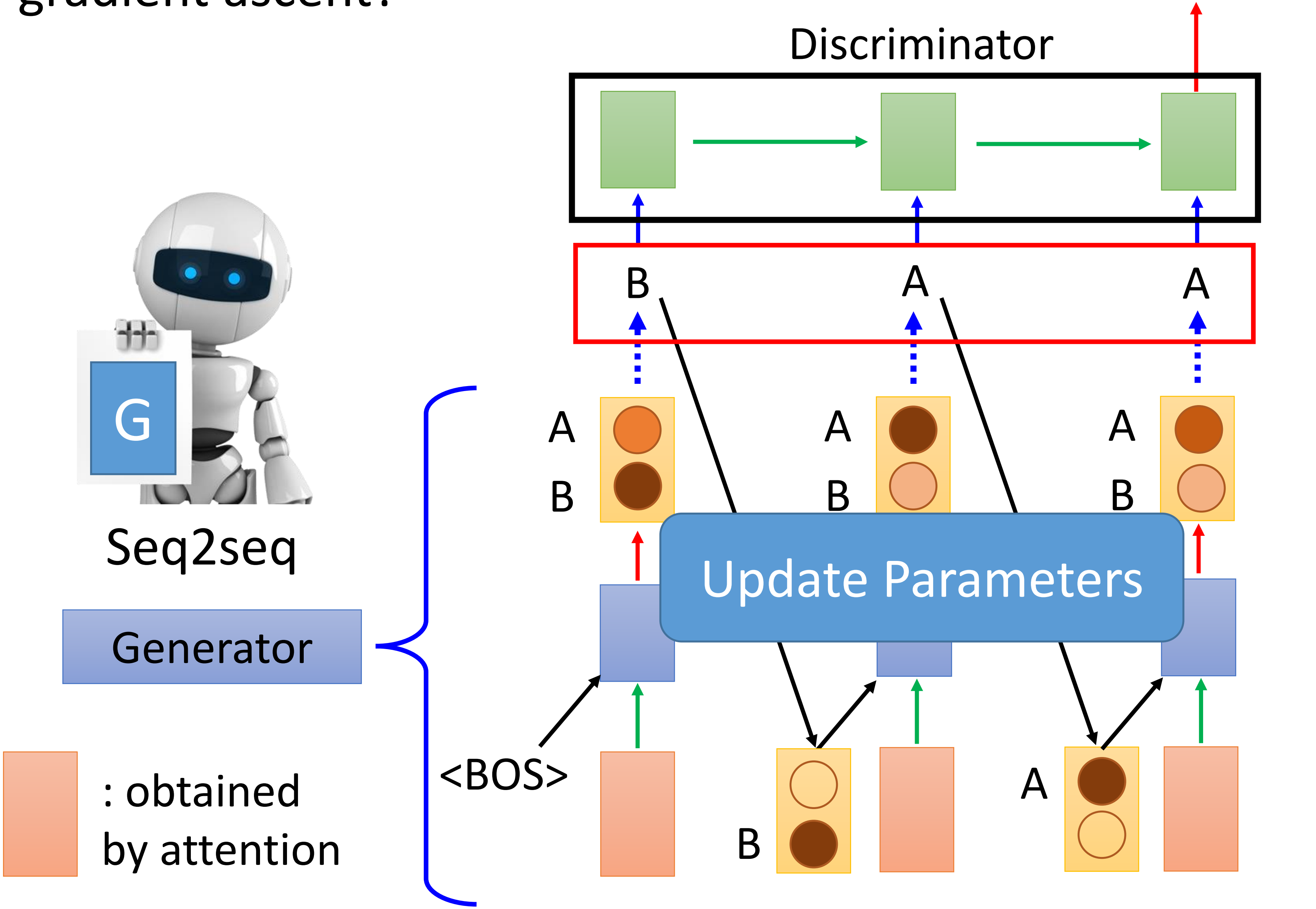
你真聰明
???????
(positive)

Seq2seq

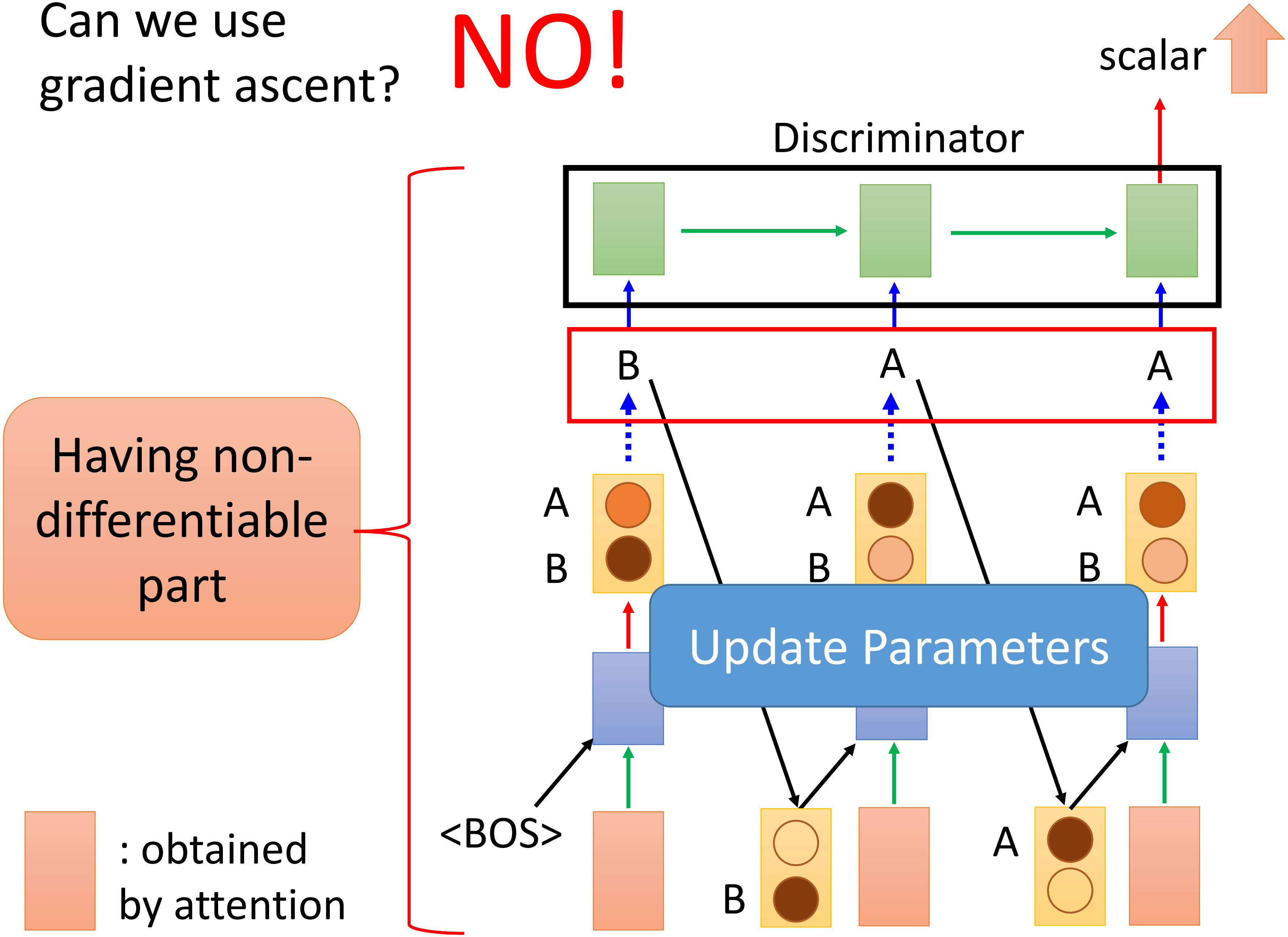
Text Style Transfer



Can we use gradient ascent? NO!



Can we use gradient ascent? **NO!**



Three Categories of Solutions

Gumbel-softmax

- [Matt J. Kusner, et al., arXiv, 2016][Weili Nie, et al. ICLR, 2019]

Continuous Input for Discriminator

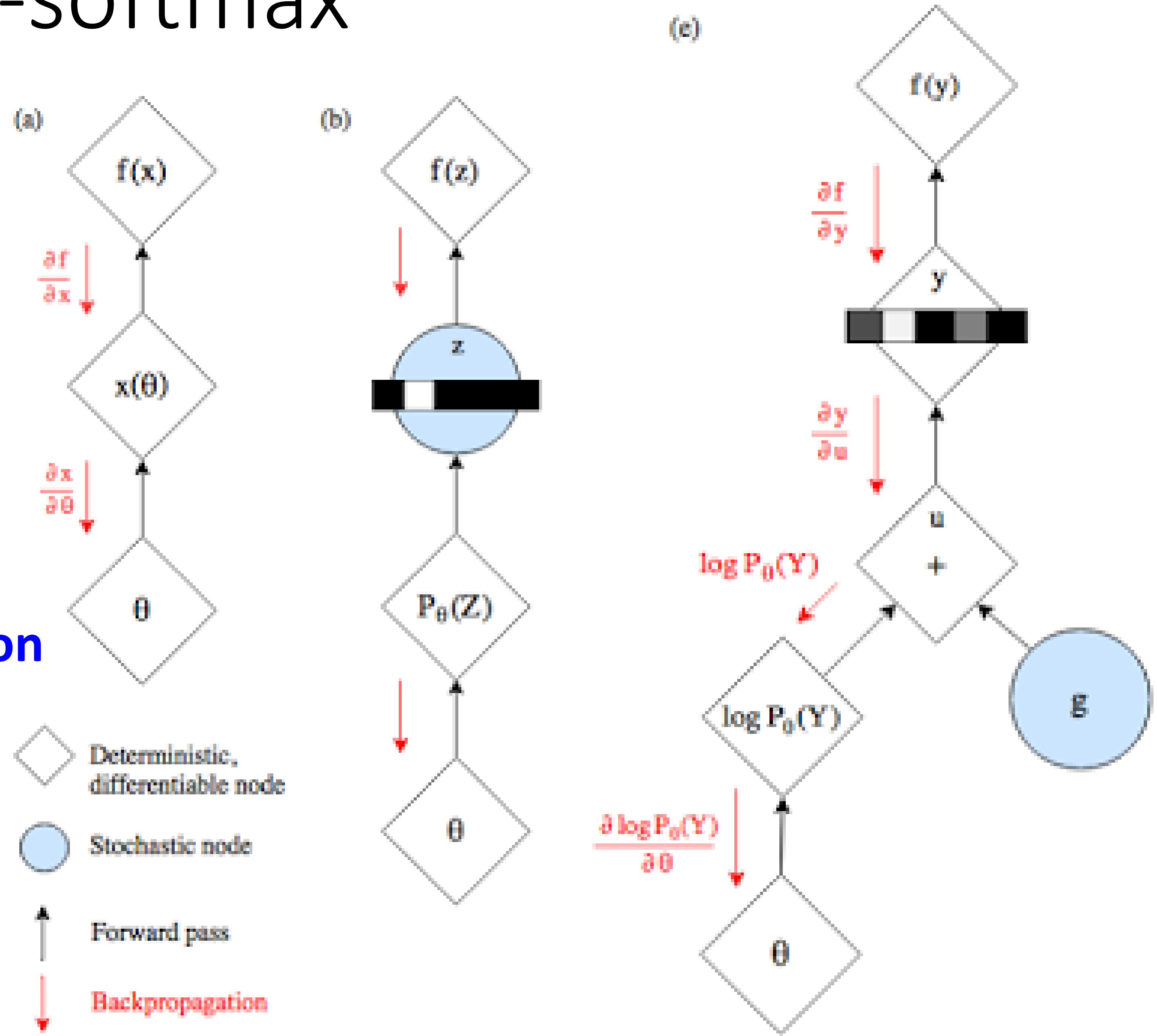
- [Sai Rajeswar, et al., arXiv, 2017][Ofir Press, et al., ICML workshop, 2017][Zhen Xu, et al., EMNLP, 2017][Alex Lamb, et al., NIPS, 2016][Yizhe Zhang, et al., ICML, 2017]

Reinforcement Learning

- [Yu, et al., AAAI, 2017][Li, et al., EMNLP, 2017][Tong Che, et al, arXiv, 2017][Jiaxian Guo, et al., AAAI, 2018][Kevin Lin, et al, NIPS, 2017][William Fedus, et al., ICLR, 2018]

Source of image: https://blog.evjang.com/2016/11/tutorial-categorical-variational.html

Gumbel-softmax



Using the **reparameterization** trick

As what people do for training **VAE**

Three Categories of Solutions

Gumbel-softmax

- [Matt J. Kusner, et al., arXiv, 2016][Weili Nie, et al. ICLR, 2019]

Continuous Input for Discriminator

- [Sai Rajeswar, et al., arXiv, 2017][Ofir Press, et al., ICML workshop, 2017][Zhen Xu, et al., EMNLP, 2017][Alex Lamb, et al., NIPS, 2016][Yizhe Zhang, et al., ICML, 2017]

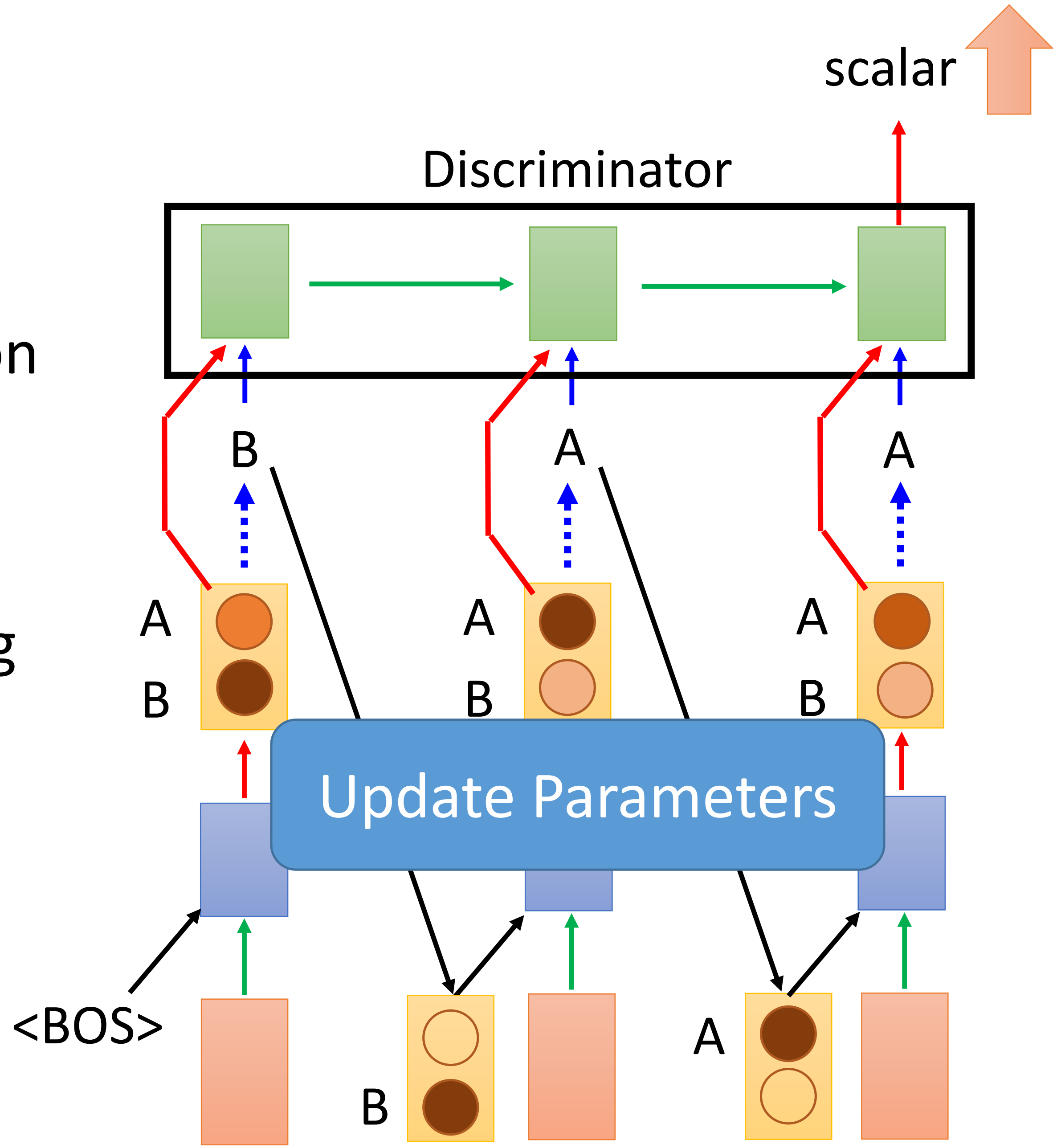
Reinforcement Learning

- [Yu, et al., AAAI, 2017][Li, et al., EMNLP, 2017][Tong Che, et al, arXiv, 2017][Jiaxian Guo, et al., AAAI, 2018][Kevin Lin, et al, NIPS, 2017][William Fedus, et al., ICLR, 2018]

Use the distribution as the input of discriminator

Avoid the sampling process

We can do backpropagation now.



What is the problem?

Discriminator with constraint (e.g. WGAN) can be helpful.

- Real sentence

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

- Generated

0.9	0.1	0.1	0	0
0.1	0.9	0.1	0	0
0	0	0.7	0.1	0
0	0	0.1	0.8	0.1
0	0	0	0.1	0.9

Can never be 1-hot

Discriminator can immediately find the difference.

Three Categories of Solutions

Gumbel-softmax

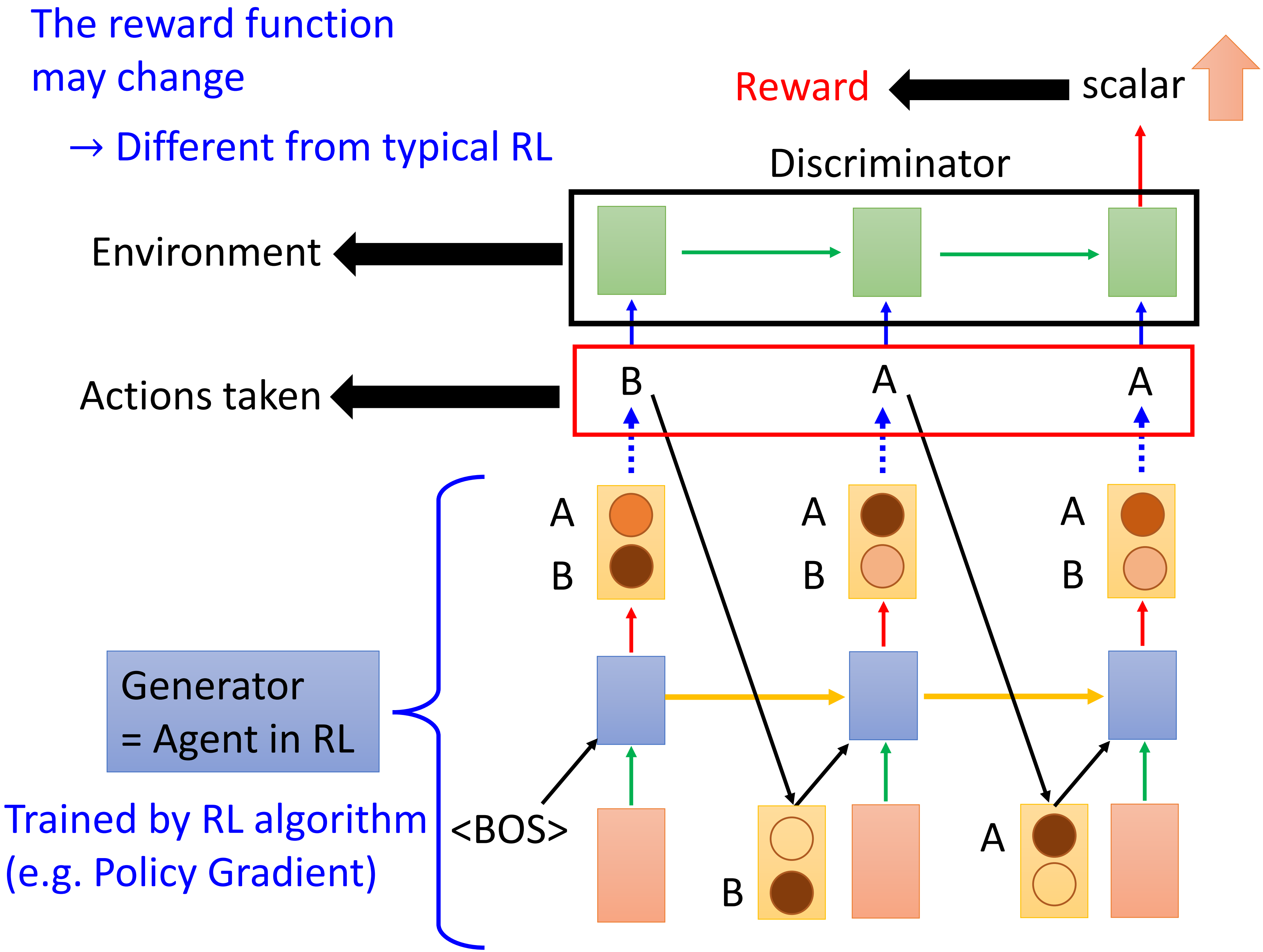
- [Matt J. Kusner, et al., arXiv, 2016][Weili Nie, et al. ICLR, 2019]

Continuous Input for Discriminator

- [Sai Rajeswar, et al., arXiv, 2017][Ofir Press, et al., ICML workshop, 2017][Zhen Xu, et al., EMNLP, 2017][Alex Lamb, et al., NIPS, 2016][Yizhe Zhang, et al., ICML, 2017]

Reinforcement Learning

- [Yu, et al., AAAI, 2017][Li, et al., EMNLP, 2017][Tong Che, et al, arXiv, 2017][Jiaxian Guo, et al., AAAI, 2018][Kevin Lin, et al, NIPS, 2017][William Fedus, et al., ICLR, 2018]

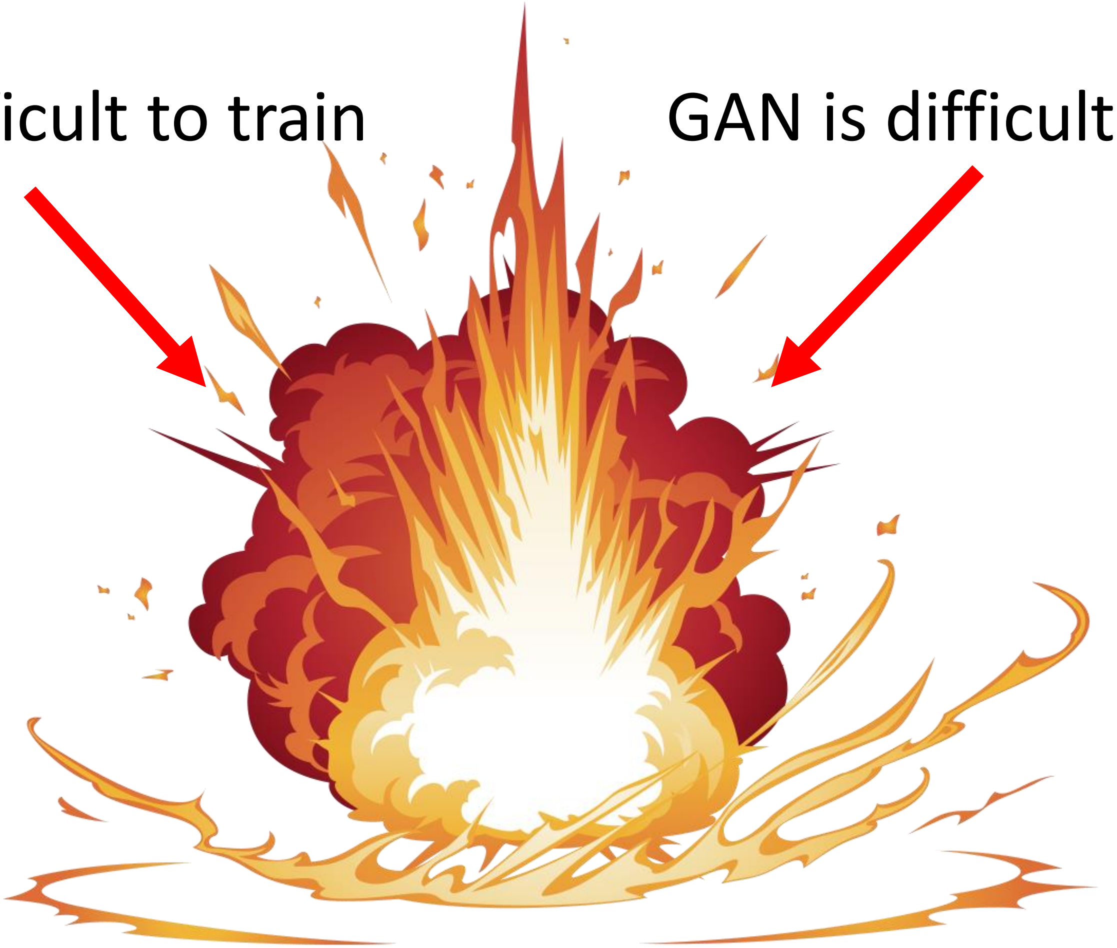


The reward function
may change
→ Different from typical RL
Reward
scalar
Discriminator
Environment
Actions taken
B
A
A
A
B
A
B
A
B
Generator
= Agent in RL
Trained by RL algorithm
(e.g. Policy Gradient)
<BOS>
B
A

Disaster ……

RL is difficult to train

GAN is difficult to train

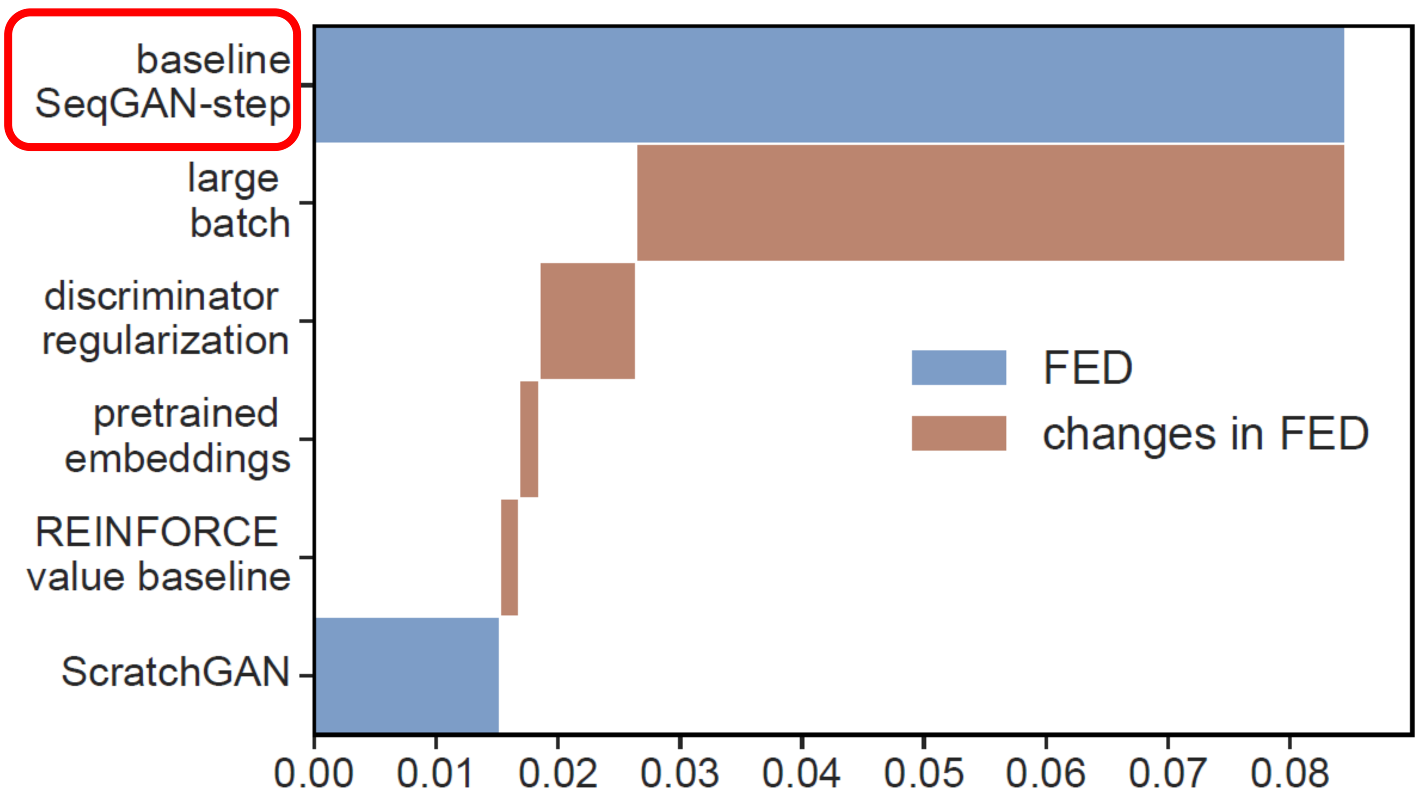


RL+GAN

Tips?

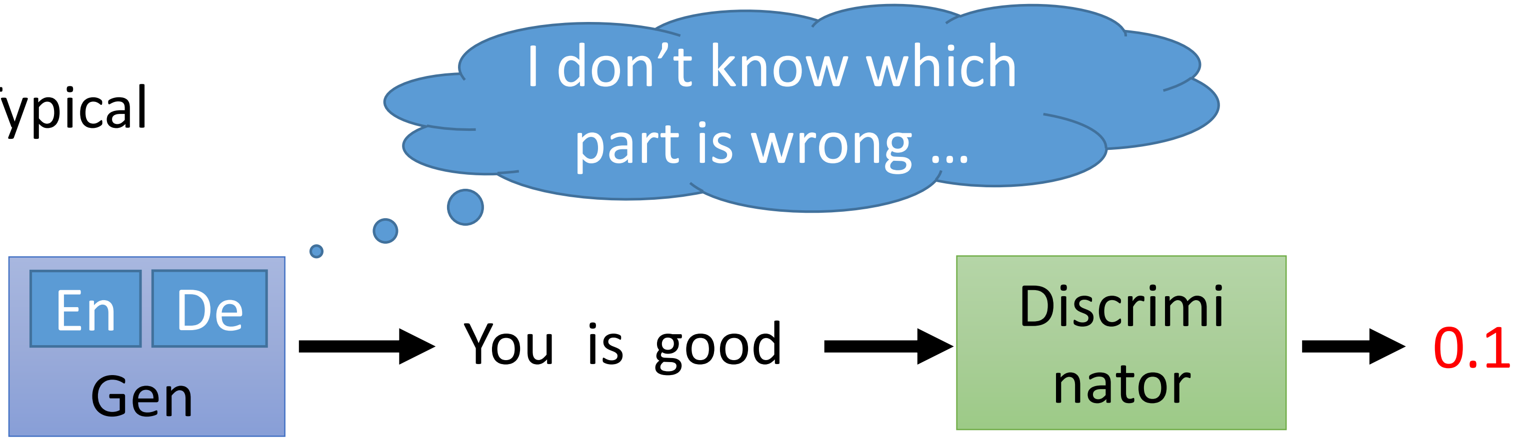
[Cyprien de Masson d'Autume, et al., arXiv 2019]

- ScratchGAN

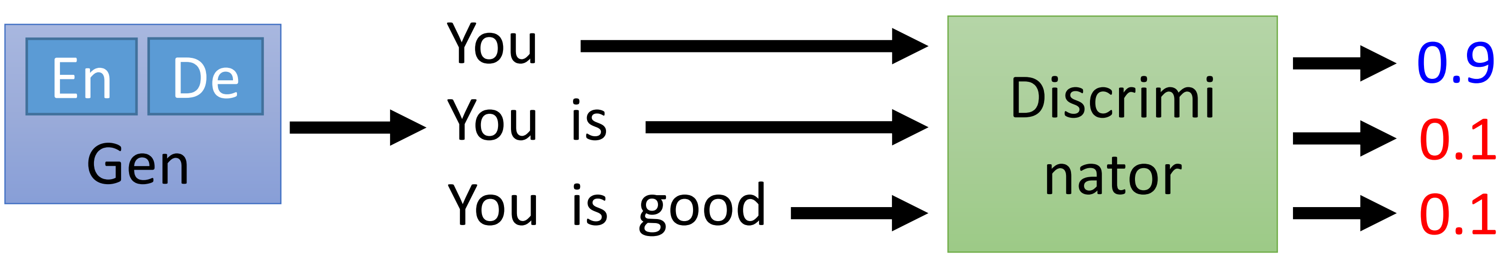


Tips?

- Typical

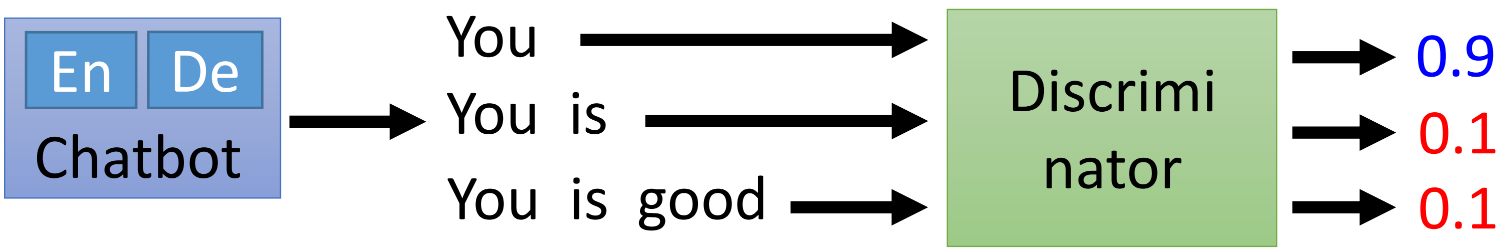


- Reward for Every Generation Step



Tips?

- Reward for Every Generation Step

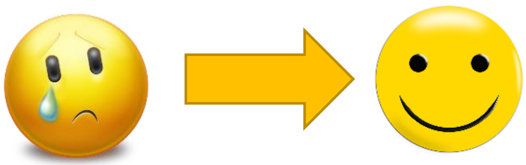


Method 1. Monte Carlo (MC) Search [Yu, et al., AAAI, 2017]

Method 2. Discriminator For Partially Decoded Sequences [Li, et al., EMNLP, 2017]

Method 3. Step-wise evaluation [Tual, Lee, TASLP, 2019][Xu, et al., EMNLP, 2018][William Fedus, et al., ICLR, 2018]

感謝 張瓊之 同學提供實驗結果



Text Style Transfer

[Lee, et al., ICASSP’18]

- From negative sentence to positive one

胃疼 , 沒睡醒 , 各種不舒服

我都想去上班了, 真夠賤的!

暈死了, 吃燒烤、竟然遇到個變態狂

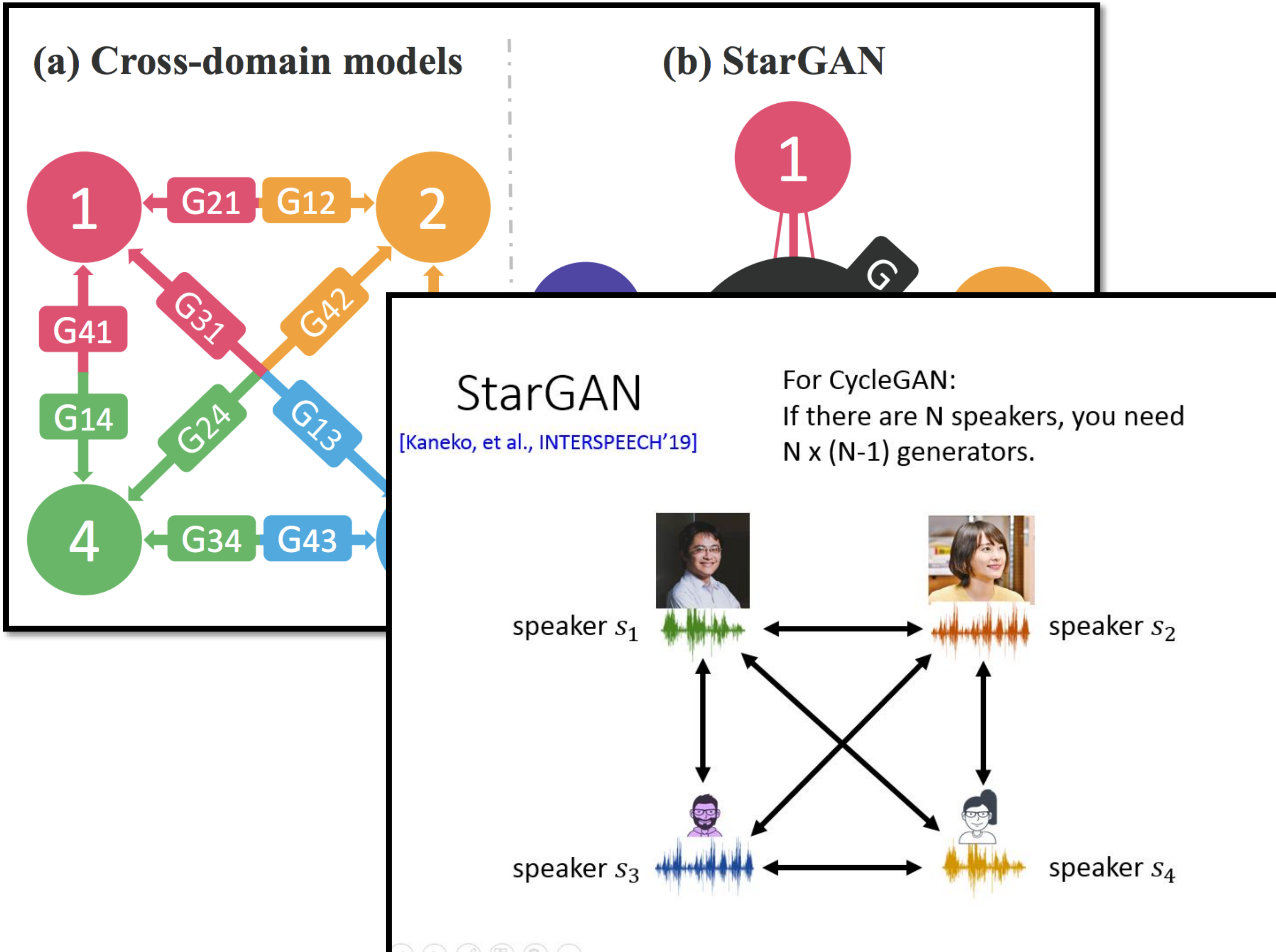
我肚子痛的厲害

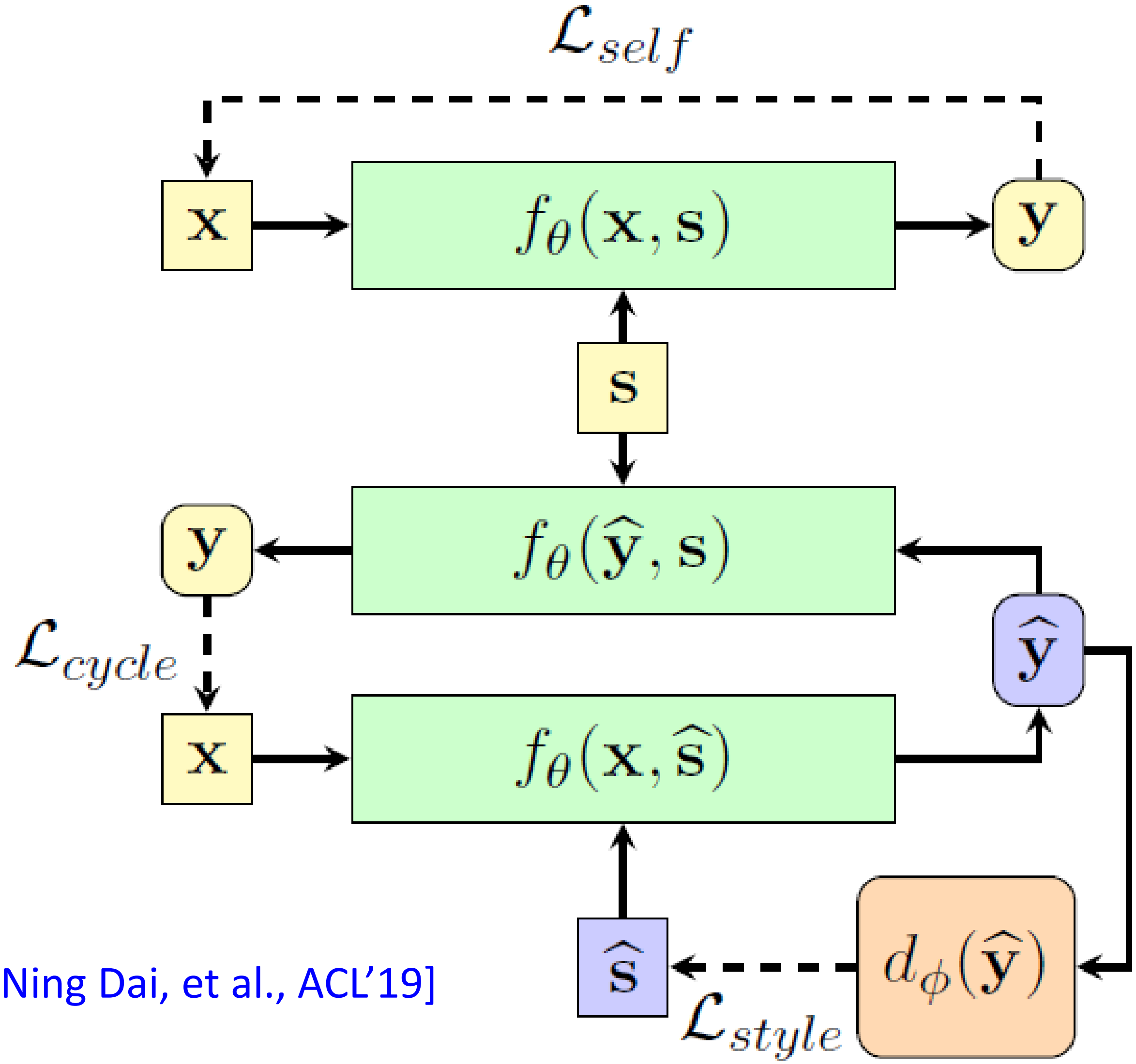
Relaxed ↔ Annoyed	
Relaxed	Sitting by the Christmas tree and watching Star Wars after cooking dinner. What a nice night 💗 🎄 ✨
Annoyed	Sitting by the computer and watching The Voice for the second time tonight. What a horrible way to start the weekend 😡 😡 😡
Annoyed	Getting a speeding ticket 50 feet in front of work is not how I wanted to start this month 😒
Relaxed	Getting a haircut followed by a cold foot massage in the morning is how I wanted to start this month 😊
Male ↔ Female	
Male	Gotta say that beard makes you look like a Viking...
Female	Gotta say that hair makes you look like a Mermaid...
Female	Awww he's so gorgeous 😍 can't wait for a cuddle. Well done 😘 xxx
Male	Bro he's so f***ing dope can't wait for a cuddle. Well done bro
Age 18-24 ↔ 65+	
18-24	You cheated on me but now I know nothing about loyalty 😂 ok
65+	You cheated on America but now I know nothing about patriotism. So ok.
65+	Ah! Sweet photo of the sisters. So happy to see them together today .
18-24	Ah 😂 Thankyou 💞 #sisters ❤️ happy to see them together today

[1]Note that using "gender" (or any other attribute for that matter) as a differentiating attribute between several bodies of text implies that there are indeed signatures of gender in the data. These signatures could be as innocuous as some first names like Mary being usually associated with women, or disheartening like biases and stereotypes exposed by statistical methods, (e.g., "man is to computer programmer as woman is to homemaker" (Bolukbasi et al., 2016)). We certainly do not condone those stereotypes, and on the contrary, we hope that showing that our models can uncover these biases might down the line turn them into powerful tools for researchers who study fairness and debiasing (Reddy & Knight, 2016).

Source of image: https://openreview.net/forum?id=H1g2NhC5KQ

[Lample, et al., ICLR'19]


(a) Cross-domain models
(b) StarGAN
1
2
4
G21
G12
G41
G14
G31
G42
G24
G13
G34
G43
G
StarGAN
[Kaneko, et al., INTERSPEECH'19]
For CycleGAN:
If there are N speakers, you need
N x (N-1) generators.
speaker s_1
speaker s_2
speaker s_3
speaker s_4

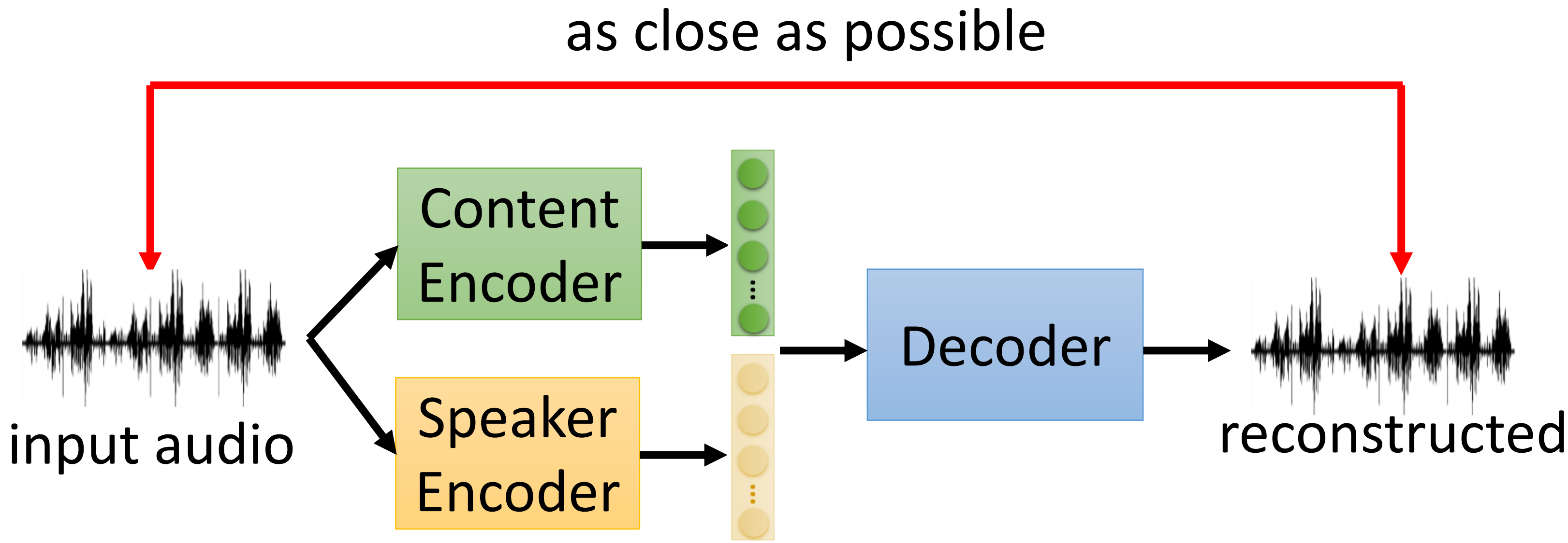


[Ning Dai, et al., ACL'19]

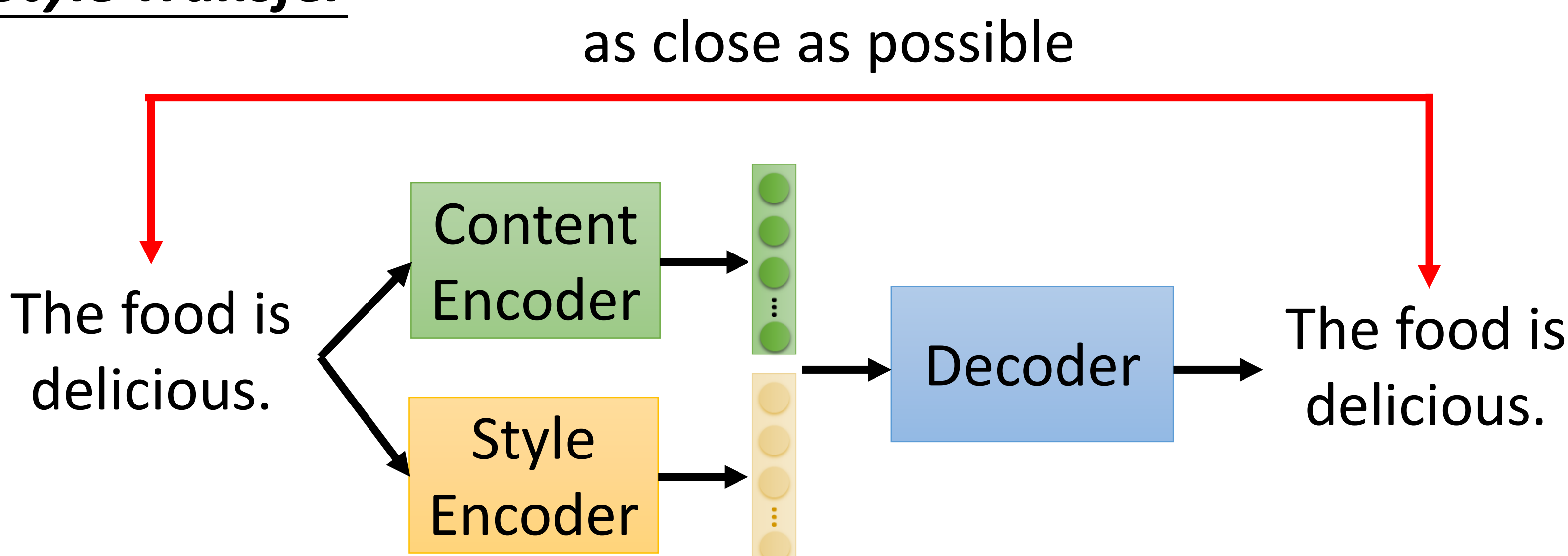
Style Transformer (Text version of StarGAN)

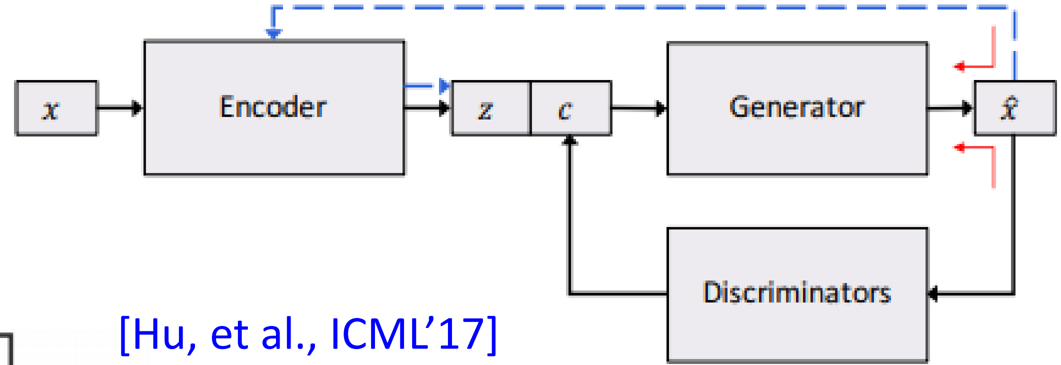
Source of image: https://arxiv.org/abs/1905.05621

Voice Conversion



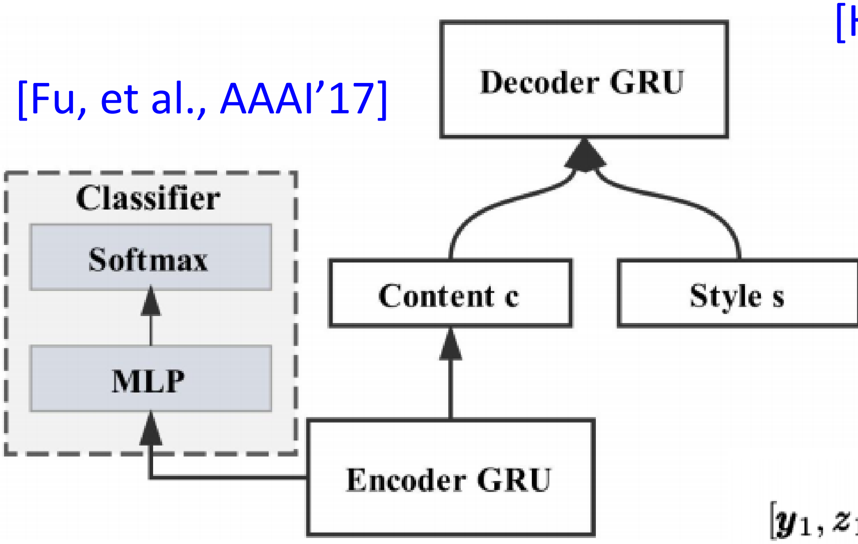
Text Style Transfer



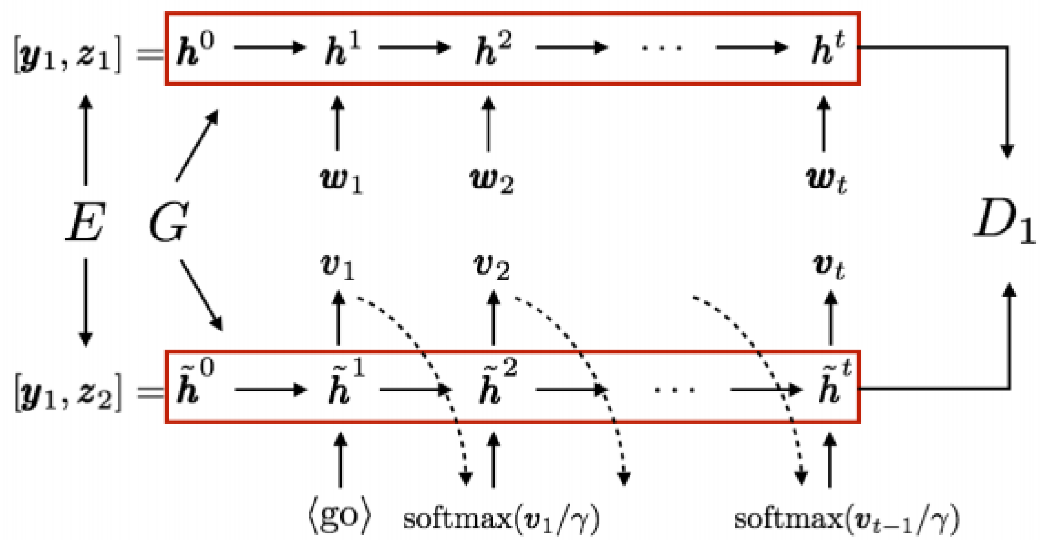


[Hu, et al., ICML'17]

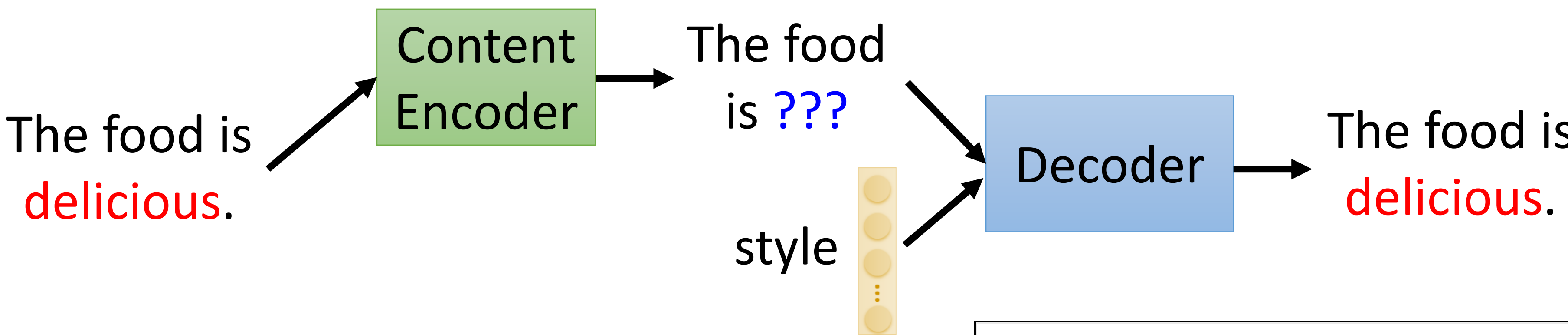
[Fu, et al., AAAI'17]

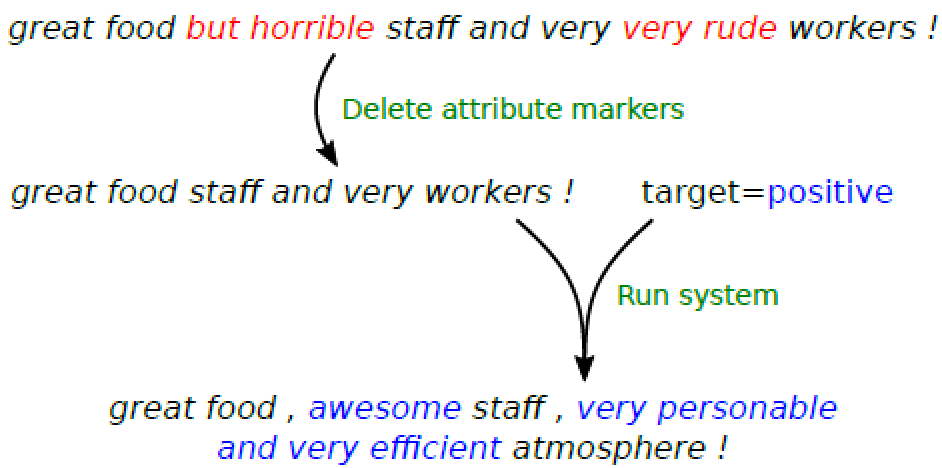


[Shen, et al., NIPS'17]

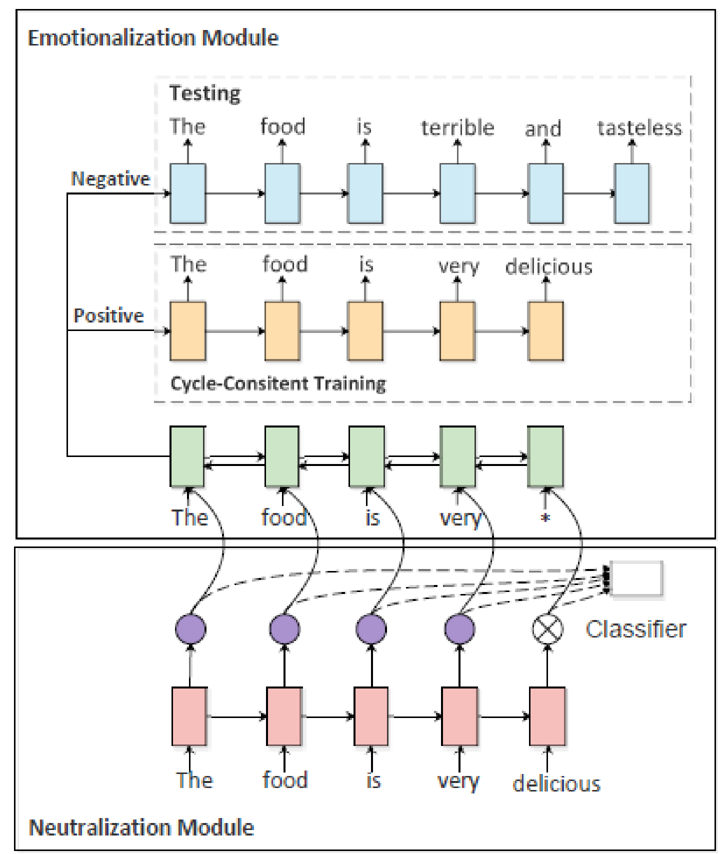


Text Style Transfer





[Li, et al., NAACL’18]



[Xu, et al., ACL’18]

Image Style

Audio Style

male

female

positive sentences

negative sentences

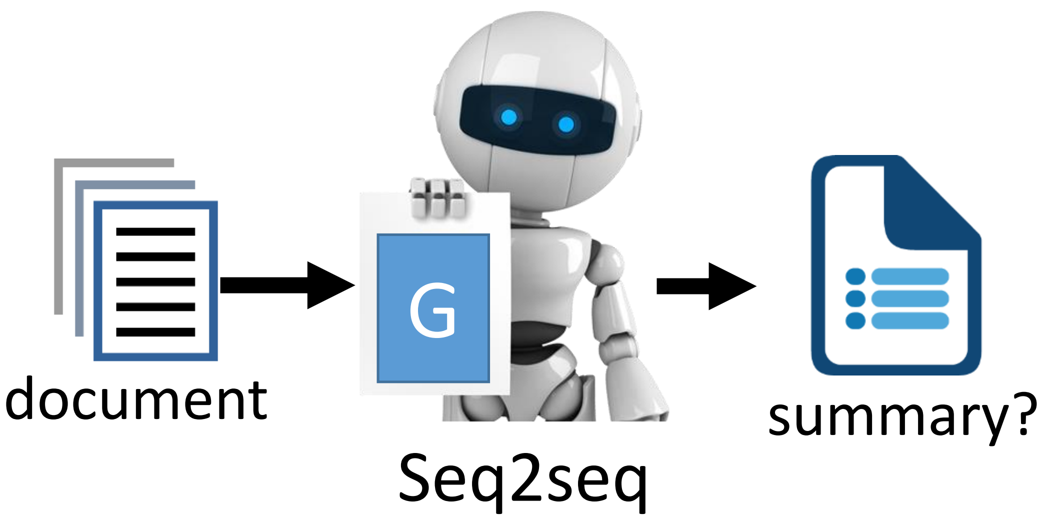
Text Style Transfer

document

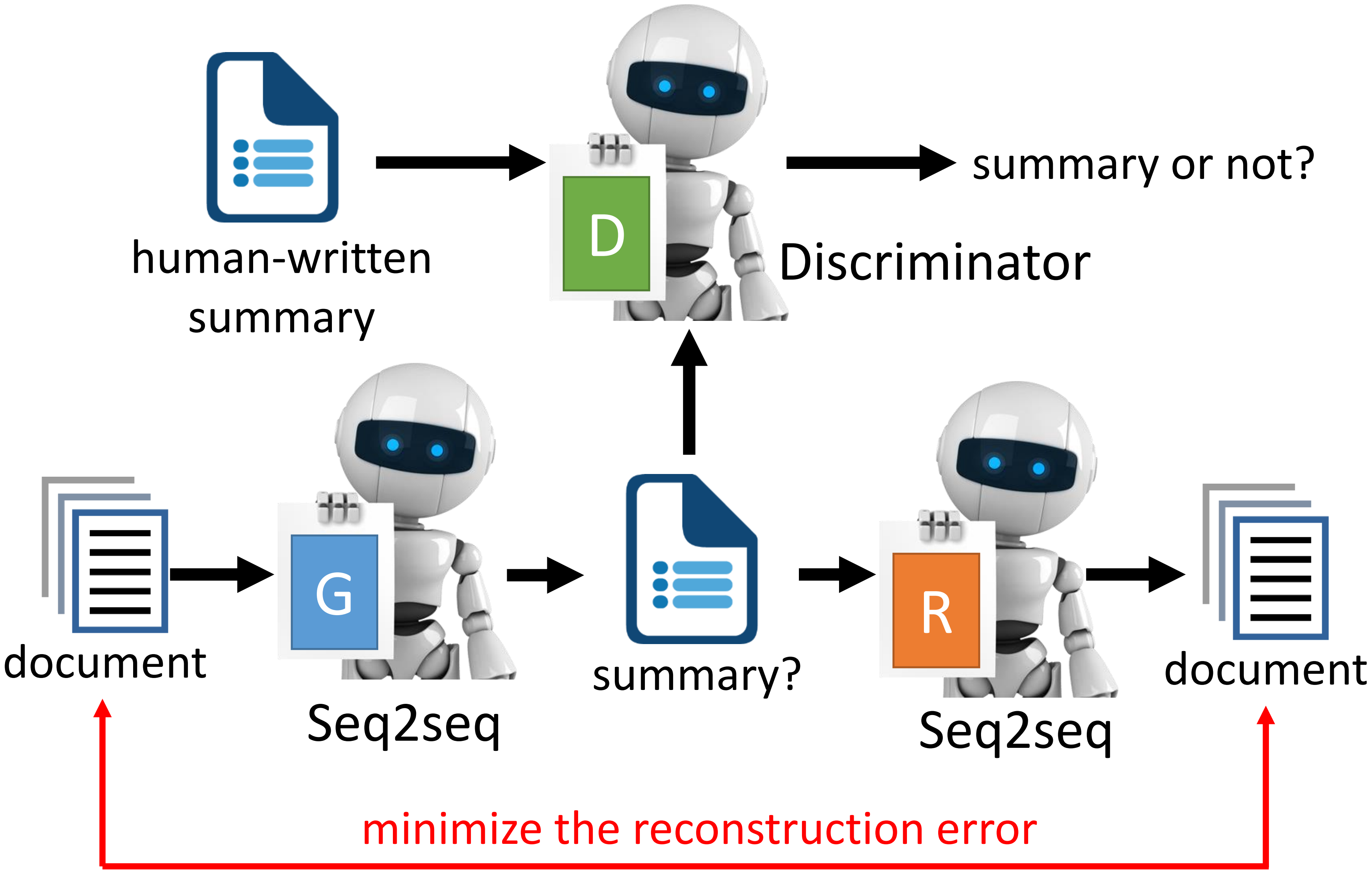
summary

Unsupervised Abstractive Summarization

Unsupervised Abstractive Summarization



Unsupervised Abstractive Summarization



Summarization

[Wang, Lee, EMNLP 2018]

English Gigaword (Document title as summary)

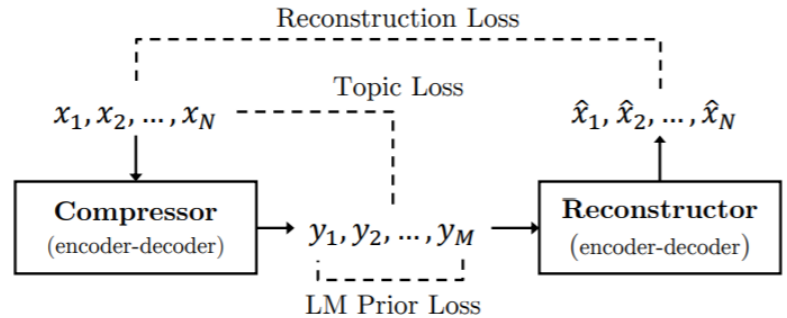
	ROUGE-1	ROUGE-2	ROUGE-L
Supervised	33.2	14.2	30.5
Trivial	21.9	7.7	20.5
Unsupervised (matched data)	28.1	10.0	25.4
Unsupervised (no matched data)	27.2	9.1	24.1

- Matched data: using the title of English Gigaword to train Discriminator
- No matched data: using the title of CNN/Diary Mail to train Discriminator

More Unsupervised Summarization

- Unsupervised summarization with language prior

[Baziotis, et al., NAACL 2019]



- Unsupervised multi-document summarization

[Chu, et al., ICML 2019]

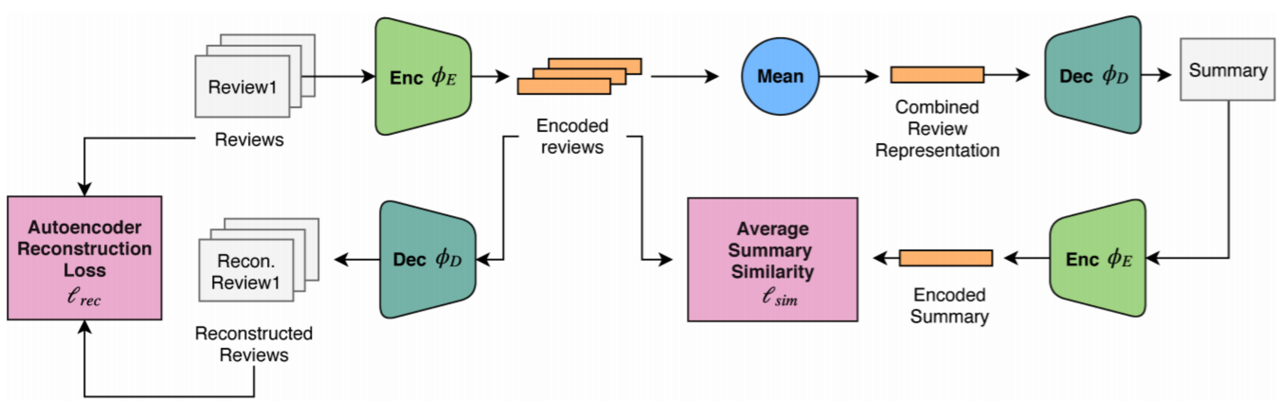


Image Style

Audio Style

male

female

positive sentences

negative sentences

Text Style Transfer

document

summary

Unsupervised Abstractive Summarization

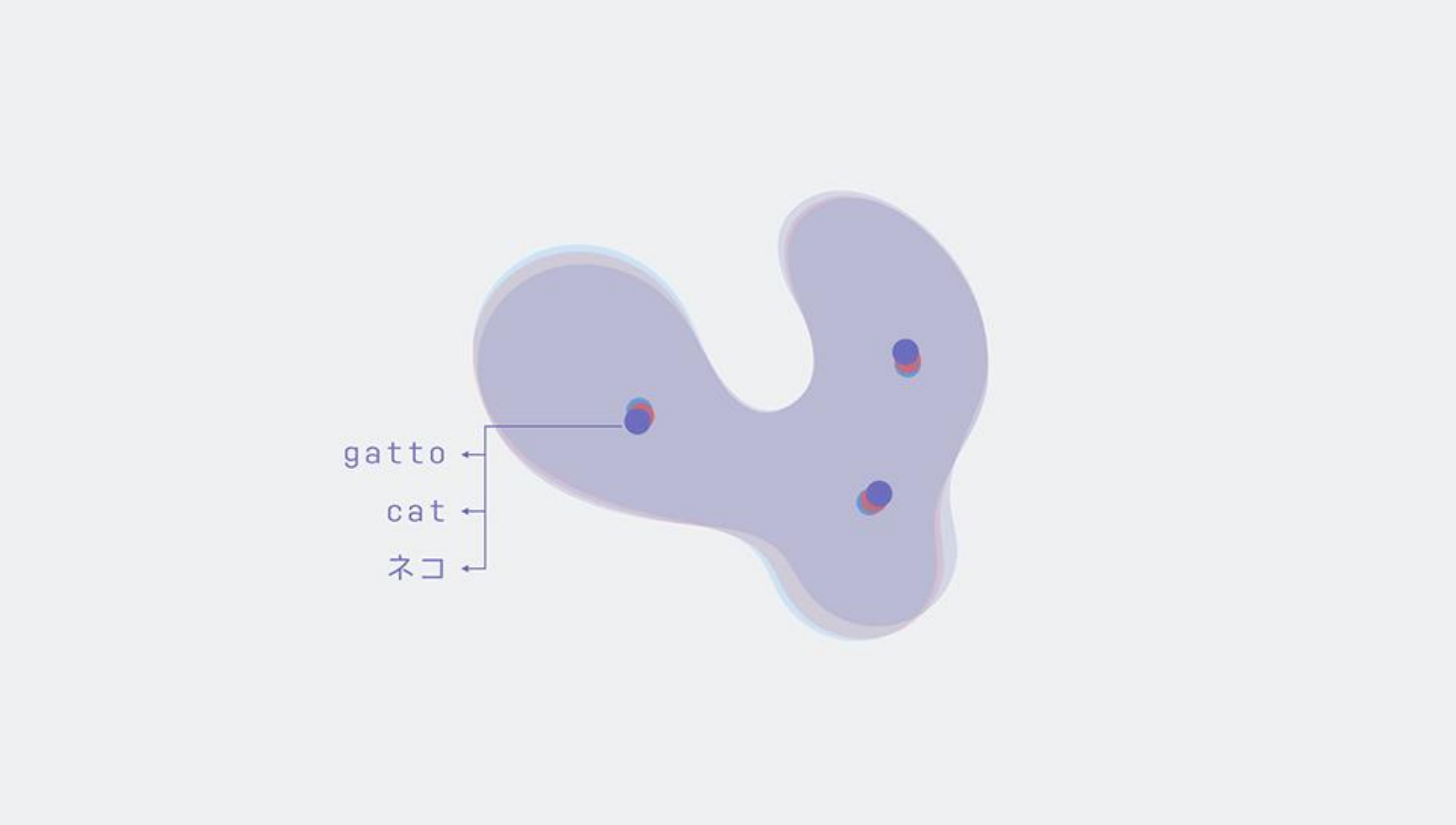
Language 1

Language 2

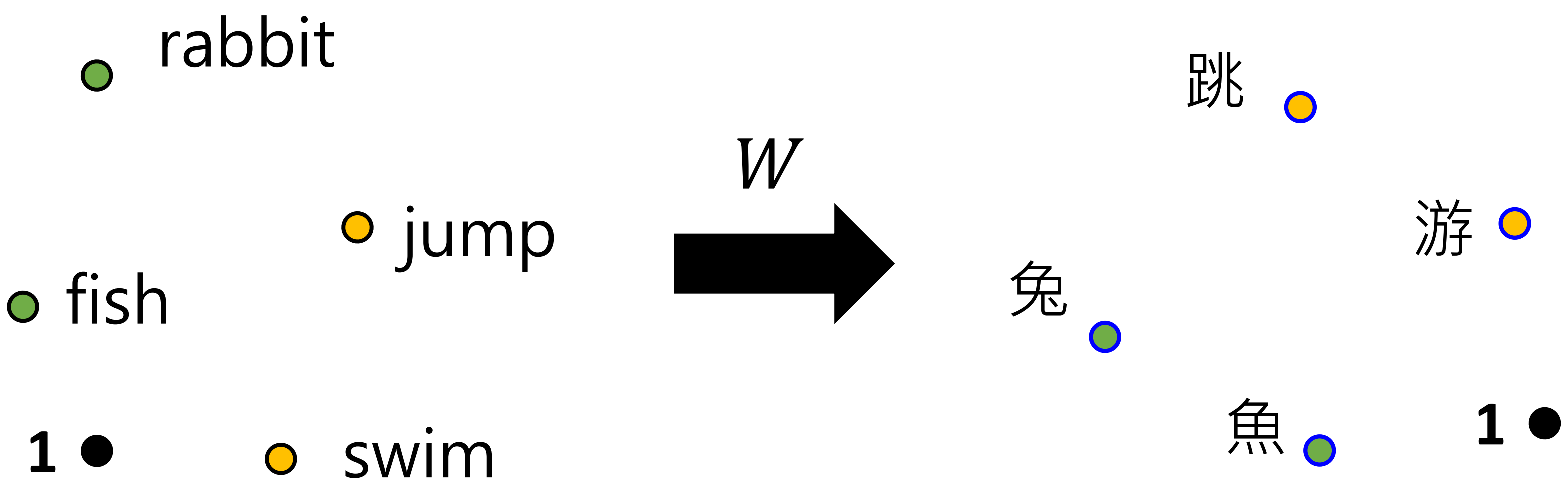
Unsupervised Translation

https://engineering.fb.com/ai-research/unsupervised-machine-translation-a-novel-approach-to-provide-fast-accurate-translations-for-more-languages/

Mapping of Word Embedding



Mapping of Word Embedding



Authors	Unsupervised step	Supervised step	Extras
Barone (2016)	GAN	None	
Zhang et al. (2017)	Wasserstein GAN	Procrustes	
Conneau et al. (2018)	GAN	Procrustes	
Hoshen and Wolf (2018)	ICP	Procrustes	Restarts
Alvarez-Melis and Jaakkola (2018)	Gromov-Wasserstein	Procrustes	
Artetxe et al. (2018)	Gromov-Wasserstein	Stochastic	
Yang et al. (2018)	Gromov-Wasserstein	MMD	
Xu et al. (2018)	GAN	Sinkhorn	Back-translation
Grave et al. (2018)	Gold-Rangarajan	Sinkhorn	

[Hartmann, et al., NeurIPS'19]

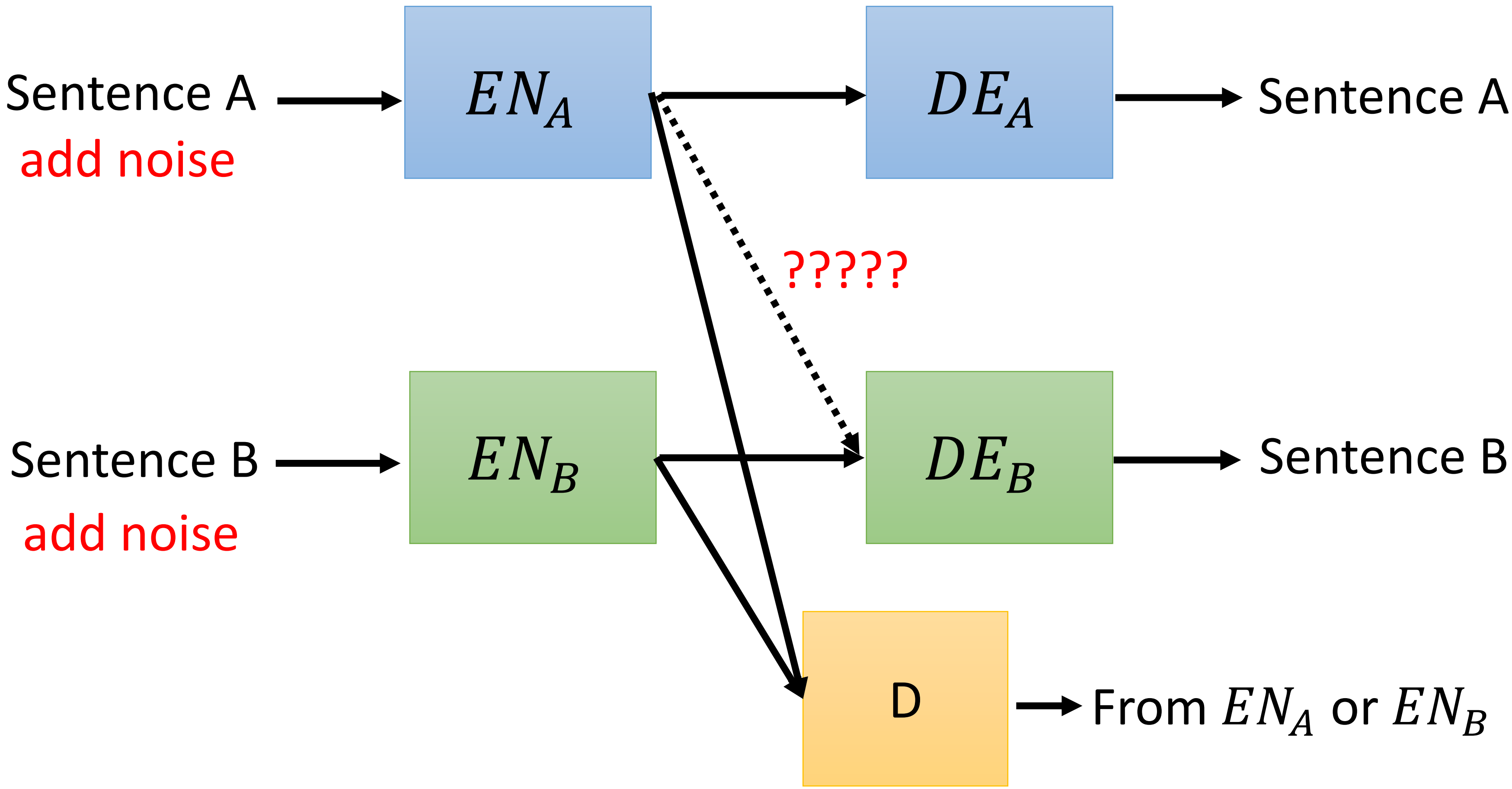
VecMap

[Artetxe, et al., ACL'18]

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)	34.93†	35.00†	25.91†	27.73†
	Faruqui and Dyer (2014)	38.40*	37.13*	27.60*	26.80*
	Shigeto et al. (2015)	41.53†	43.07†	31.04†	33.73†
	Dinu et al. (2015)	37.7	38.93*	29.14*	30.40*
	Lazaridou et al. (2015)	40.2	-	-	-
	Xing et al. (2015)	36.87†	41.27†	28.23†	31.20†
	Zhang et al. (2016)	36.73†	40.80†	28.16†	31.07†
	Artetxe et al. (2016)	39.27	41.87*	30.62*	31.40*
	Artetxe et al. (2017)	39.67	40.87	28.72	-
	Smith et al. (2017)	43.1	43.33†	29.42†	35.13†
	Artetxe et al. (2018a)	45.27	44.13	**32.94**	36.60
25 dict.	Artetxe et al. (2017)	37.27	39.60	28.16	-
Init. heurist.	Smith et al. (2017), cognates	39.9	-	-	-
	Artetxe et al. (2017), num.	39.40	40.27	26.47	-
None	Zhang et al. (2017a), $\lambda = 1$	0.00*	0.00*	0.00*	0.00*
	Zhang et al. (2017a), $\lambda = 10$	0.00*	0.00*	0.01*	0.01*
	Conneau et al. (2018), code‡	45.15*	46.83*	0.38*	35.38*
	Conneau et al. (2018), paper‡	45.1	0.01*	0.01*	35.44*
	Proposed method	**48.13**	**48.19**	32.63	**37.33**

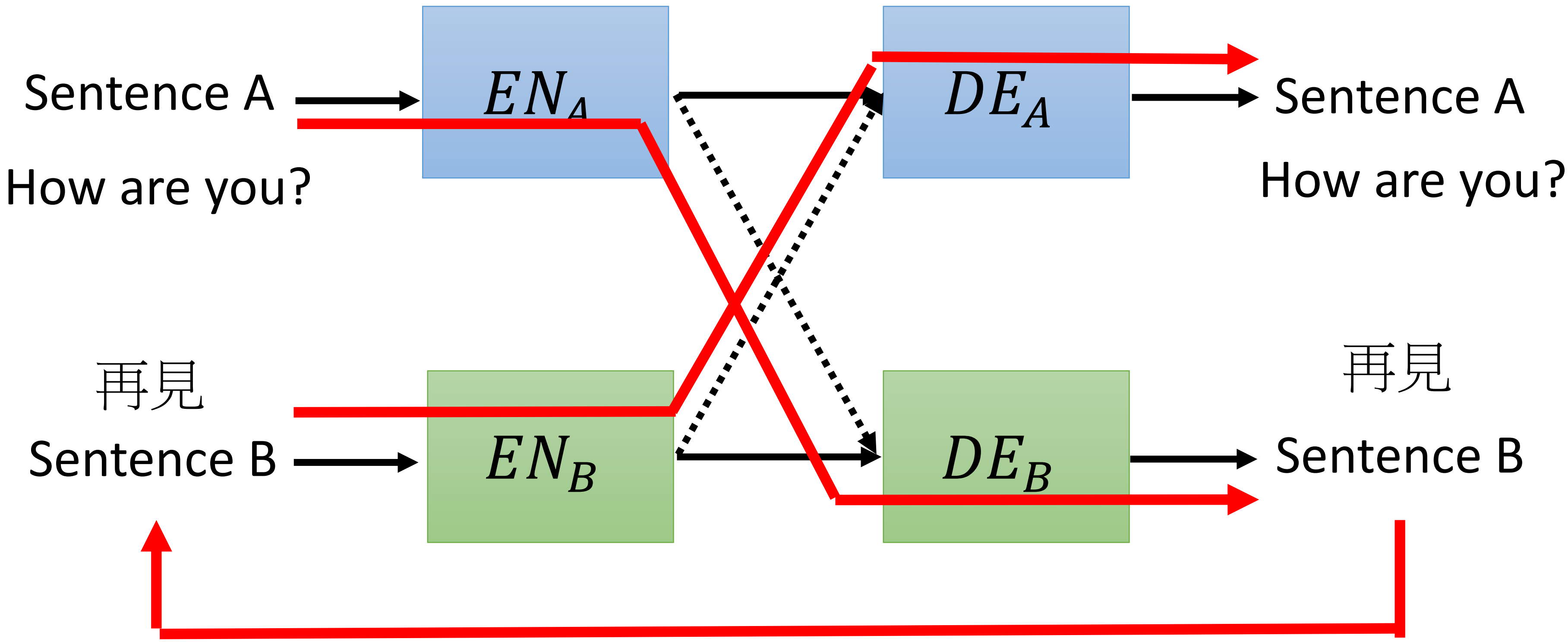
[Lample, et al., ICLR, 2018]

Unsupervised Translation

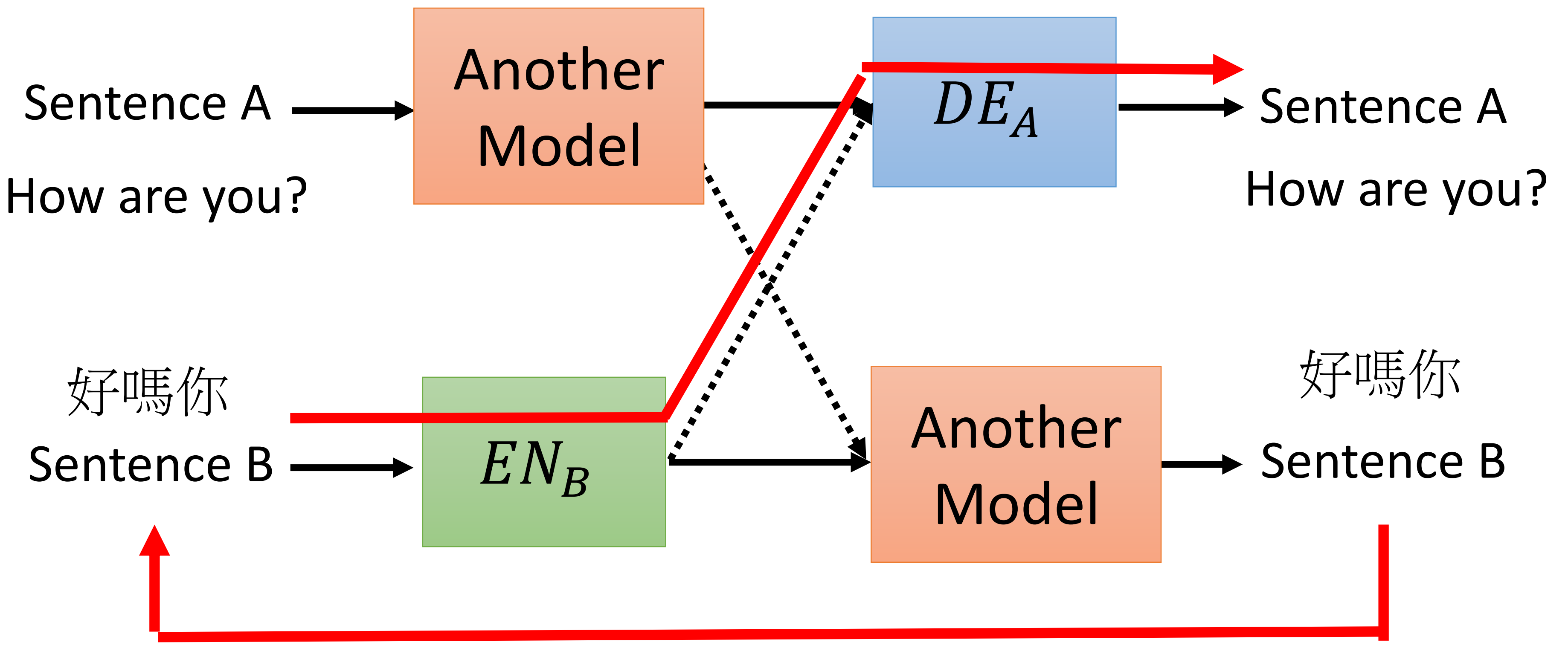


[Lample, et al., ICLR, 2018]

Unsupervised Translation

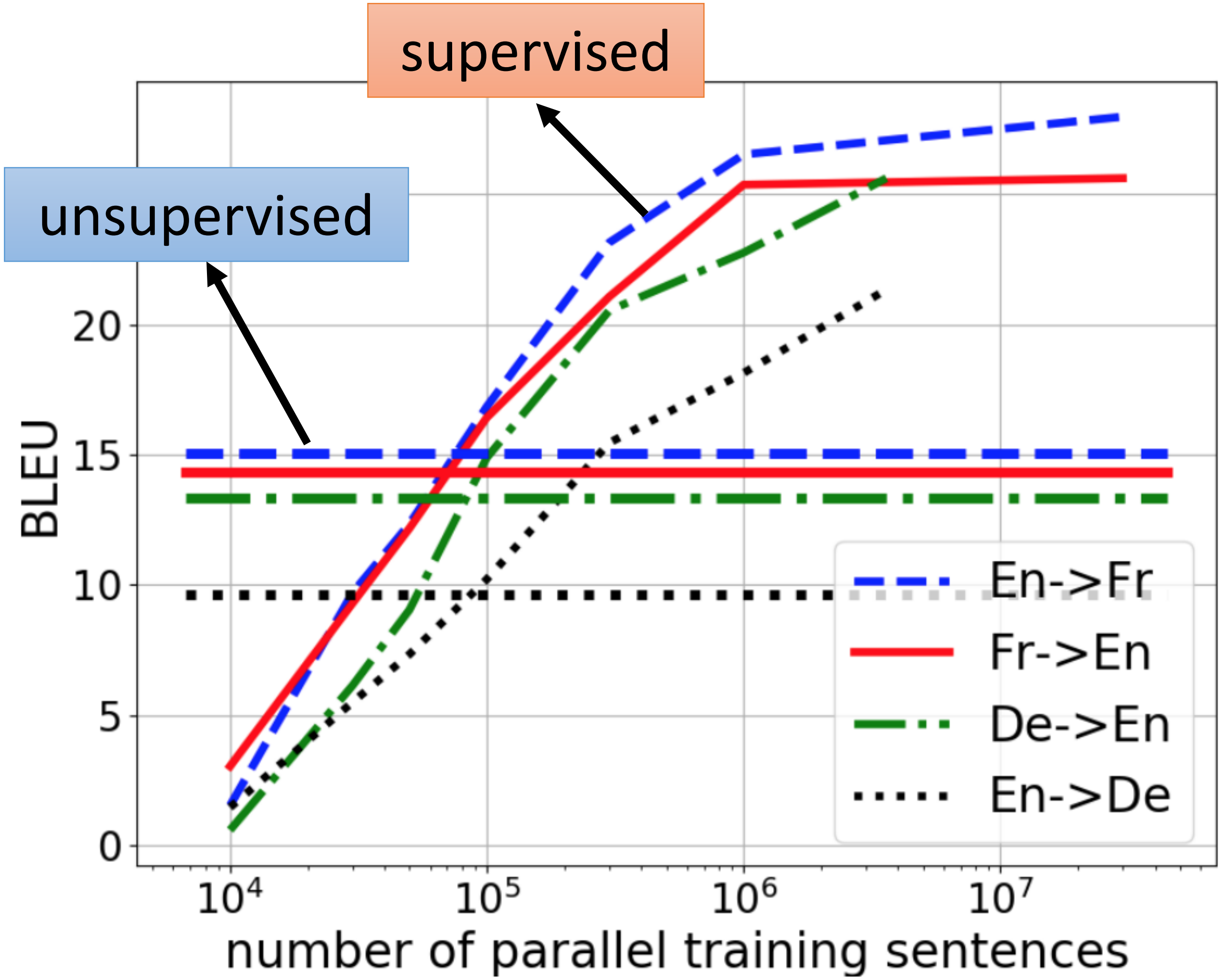


Unsupervised Translation



Start from another unsupervised translation model
(word embedding translation model)

[Lample, et al., ICLR, 2018]



Unsupervised learning with 10M sentences = **Supervised** learning with 100K sentence pairs

	en → fr	fr → en	en → de	de → en	en → ro	ro → en	en → ru	ru → en
Unsupervised PBSMT								
Unsupervised phrase table	-	17.50	-	15.63	-	14.10	-	8.08
Back-translation - Iter. 1	24.79	26.16	15.92	22.43	18.21	21.49	11.04	15.16
Back-translation - Iter. 2	27.32	26.80	17.65	22.85	20.61	22.52	12.87	16.42
Back-translation - Iter. 3	27.77	26.93	17.94	22.87	21.18	22.99	13.13	16.52
Back-translation - Iter. 4	27.84	27.20	17.77	22.68	21.33	23.01	13.37	**16.62**
Back-translation - Iter. 5	**28.11**	27.16	-	-	-	-	-	-
Unsupervised NMT								
LSTM	24.48	23.74						
Transformer	25.14	24.18						
Phrase-based + Neural network								
NMT + PBSMT	27.12	26.29						
PBSMT + NMT	27.60	**27.68**						

[Lample, et al., EMNLP'18]

		en-fr	fr-en	en-de	de-en	en-ro	ro-en
Previous state-of-the-art - Lample et al. (2018b)							
NMT		25.1	24.2	17.2	21.0	21.2	19.4
PBSMT		28.1	27.2	17.8	22.7	21.3	23.0
PBSMT + NMT		27.6	27.7	20.2	25.2	25.1	23.9
Our results for different encoder and decoder initializations							
EMB	EMB	29.4	29.4	21.3	27.3	27.5	26.6
-	-	13.0	15.8	6.7	15.3	18.9	18.3
-	CLM	25.3	26.4	19.2	26.0	25.7	24.6
-	MLM	29.2	29.1	21.6	28.6	28.2	27.3
CLM	-	28.7	28.2	24.4	30.3	29.2	28.0
CLM	CLM	30.4	30.0	22.7	30.5	29.0	27.8
CLM	MLM	32.3	31.6	24.3	32.5	31.6	29.8
MLM	-	31.6	32.1	**27.0**	33.2	31.8	30.5
MLM	CLM	**33.4**	32.3	24.9	32.9	31.7	30.4
MLM	MLM	**33.4**	**33.3**	26.4	**34.3**	**33.3**	**31.8**

[Lample, et al., NeurIPS'19]

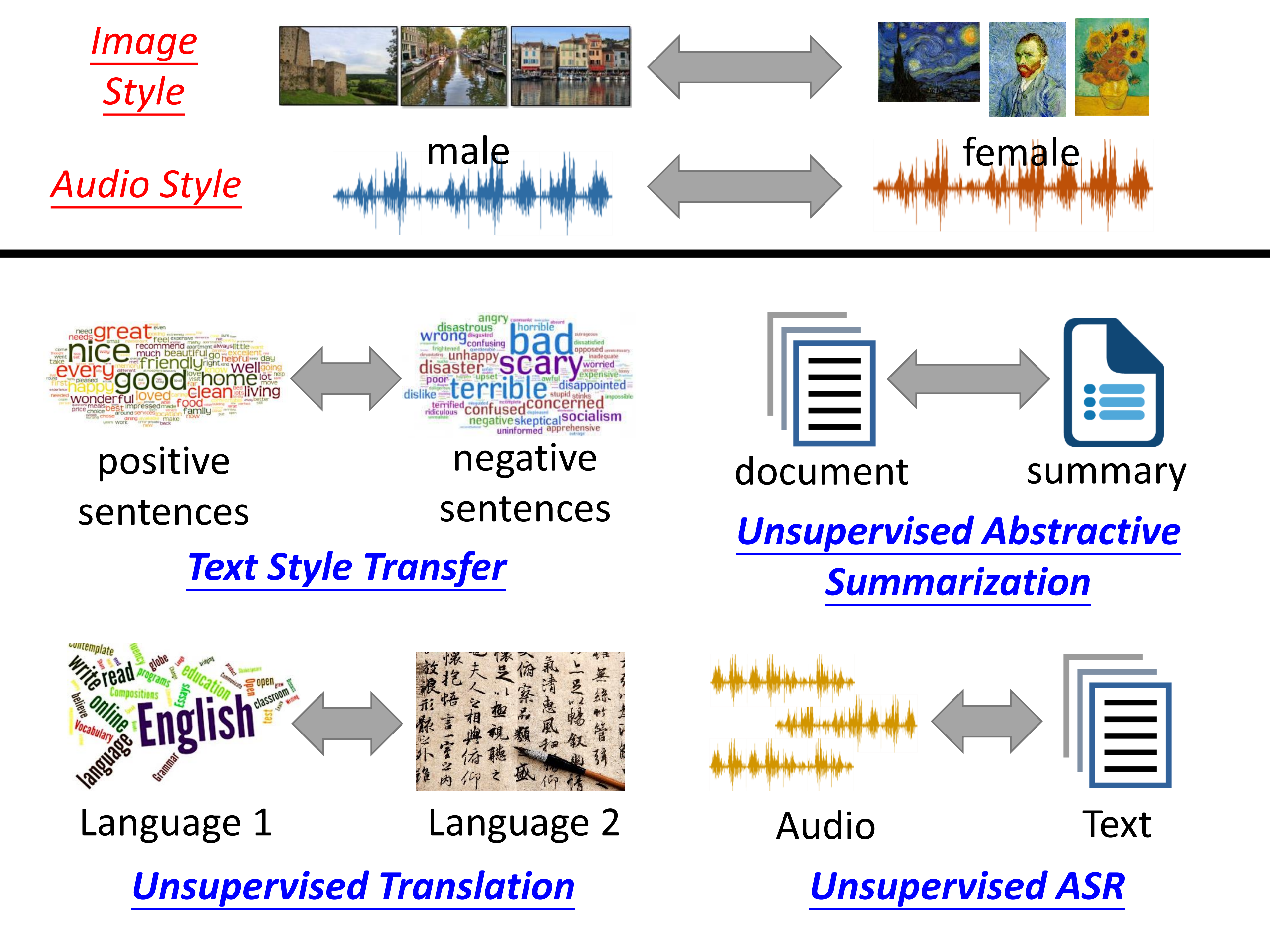
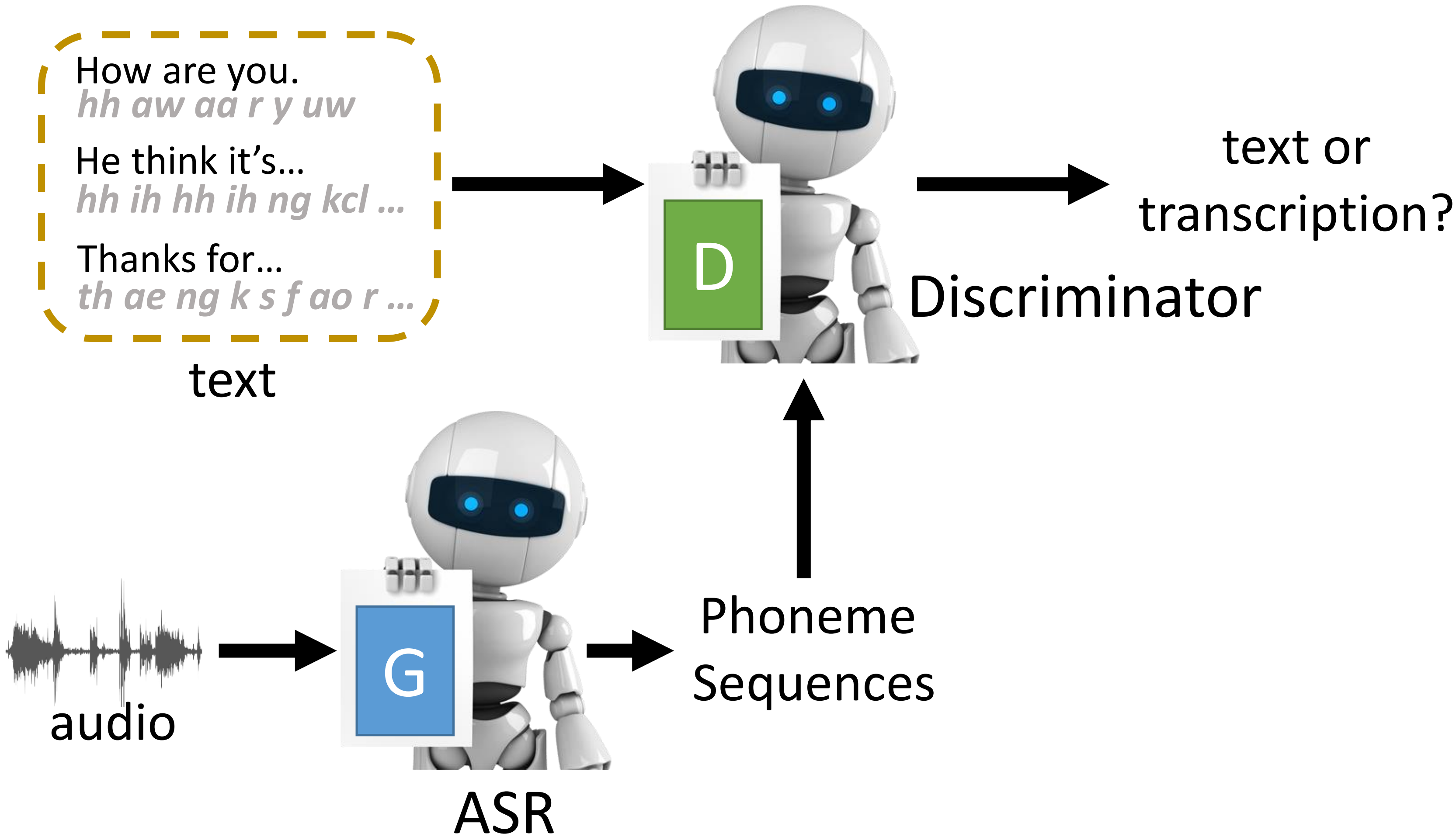
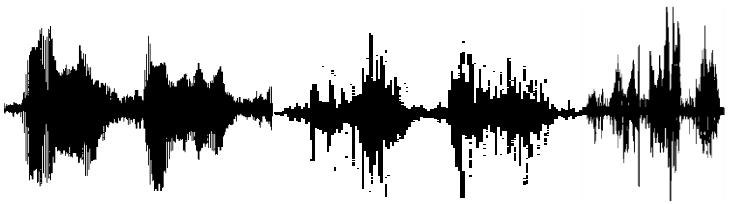


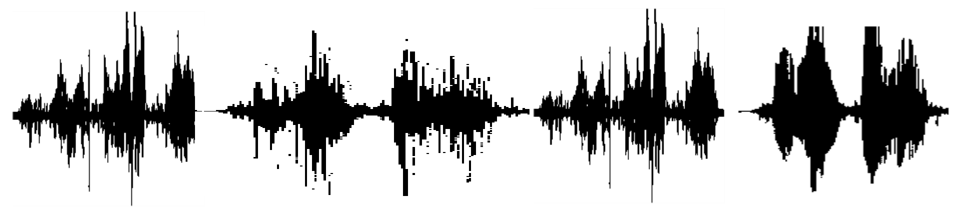
Image Style
Audio Style
male
female
positive sentences
negative sentences
Text Style Transfer
document
summary
Unsupervised Abstractive Summarization
Language 1
Language 2
Unsupervised Translation
Audio
Text
Unsupervised ASR

Unsupervised Speech Recognition

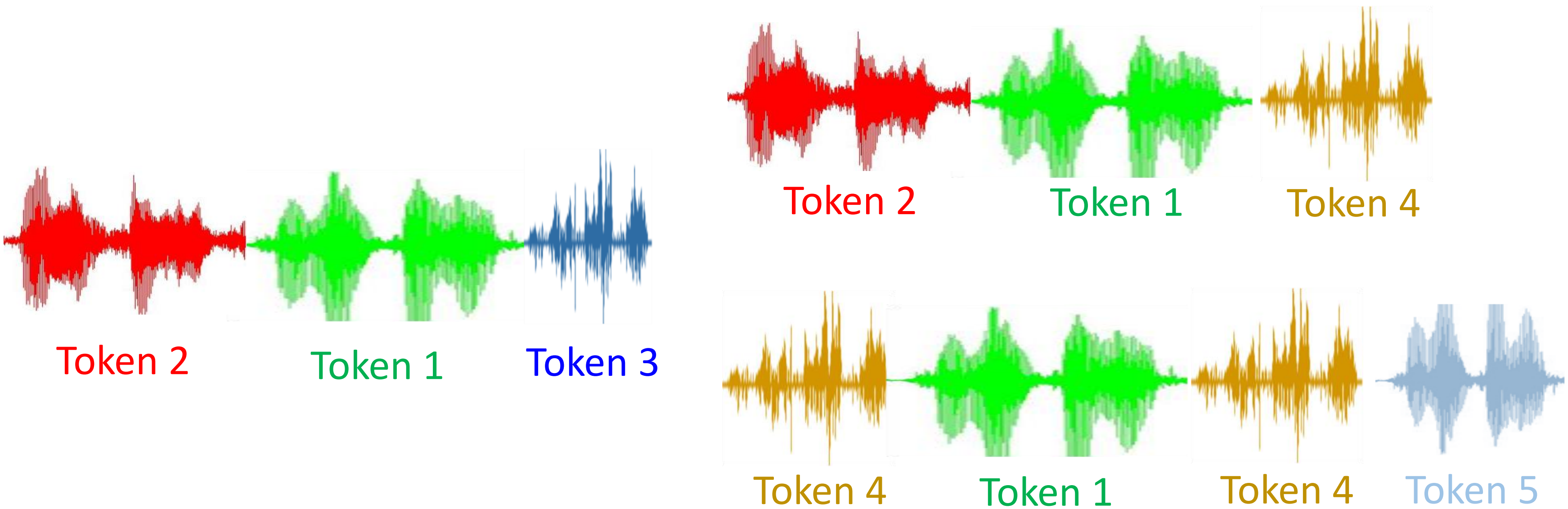


Acoustic Token Discovery





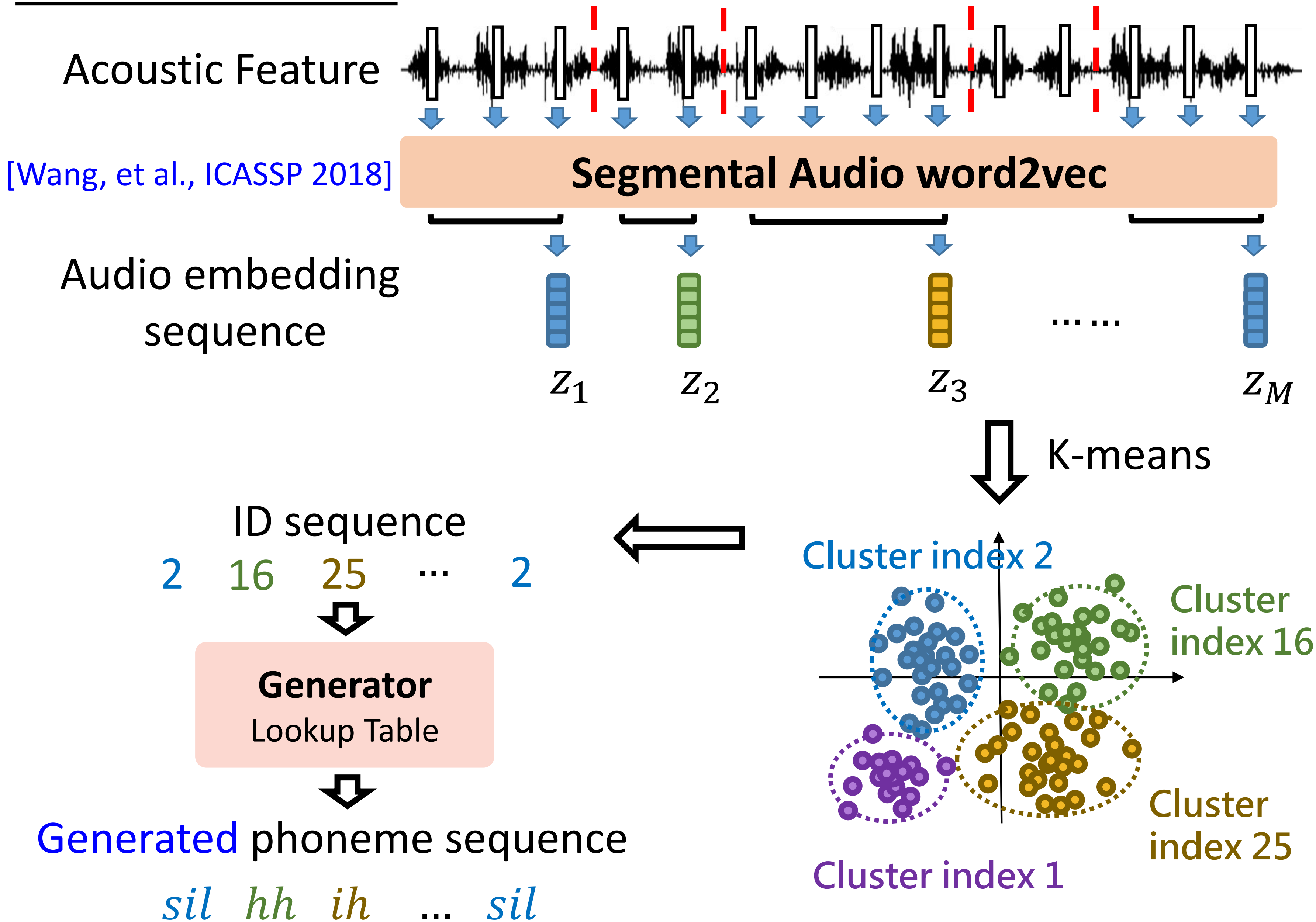
Acoustic Token Discovery



Acoustic tokens can be discovered from audio collection without text annotation.

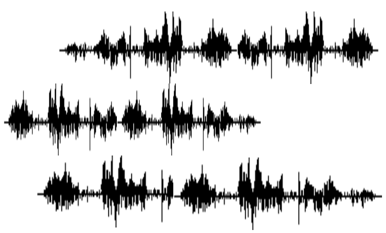
Acoustic tokens: chunks of acoustically similar audio segments with token IDs

Generator (v1)



Experiment

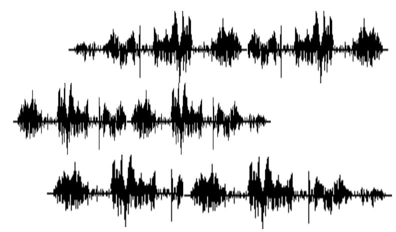
Matched Case
(Oracle)



4620
(TIMIT)

4620
(TIMIT)

Nonmatched
Case



3620
(TIMIT)

1000
(TIMIT)

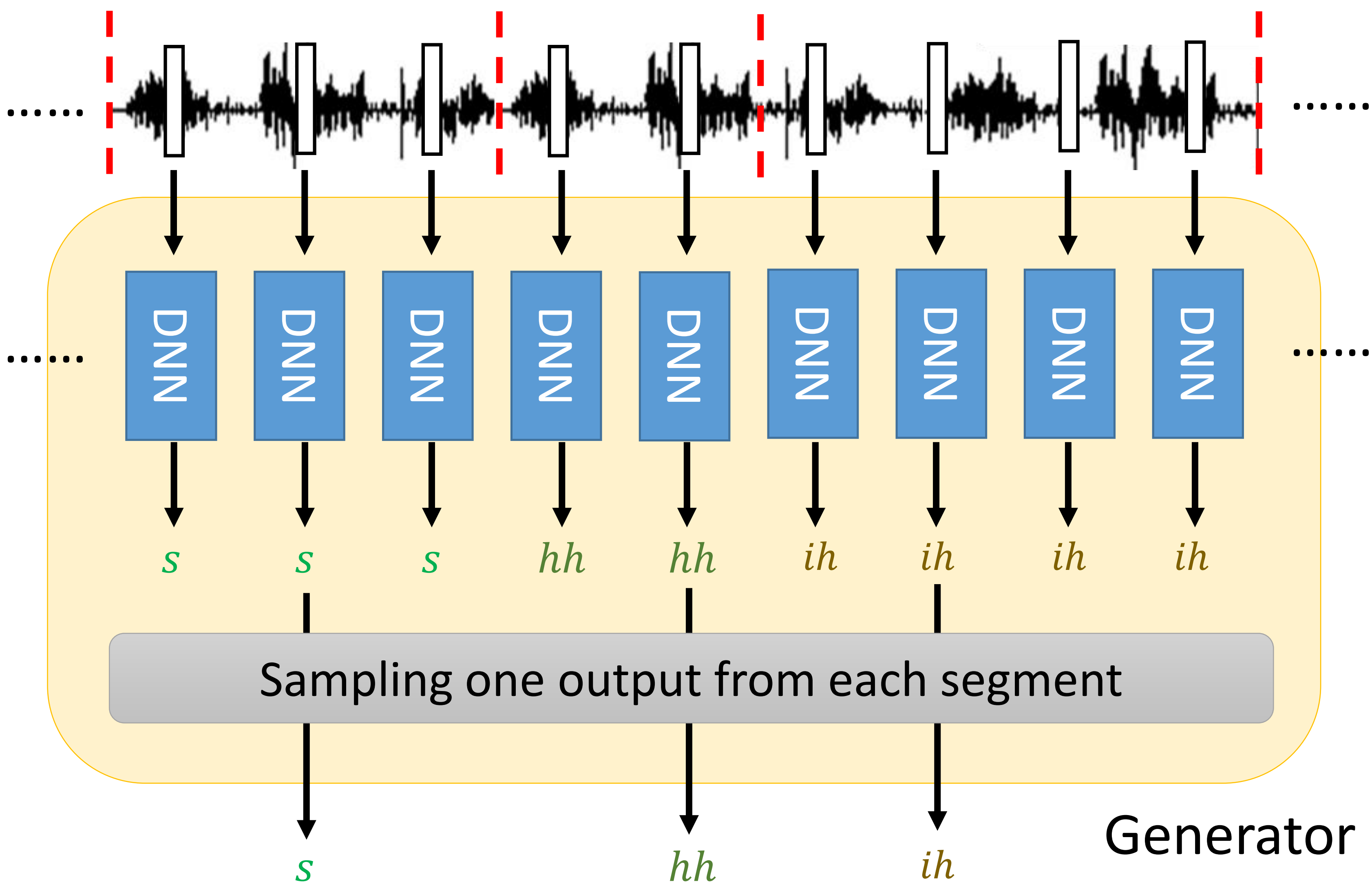
Experimental Results [Liu, et al., INTERSPEECH, 2018]

Approaches	PER	
	Matched	Nonmatched
Supervised		
RNN Transducer	17.7	-
Standard HMMs	21.5	-

Generator (v2)

Phoneme boundaries obtained by Gate Activation Signals (GAS)

[Wang, et al., INTERSPEECH 2017]

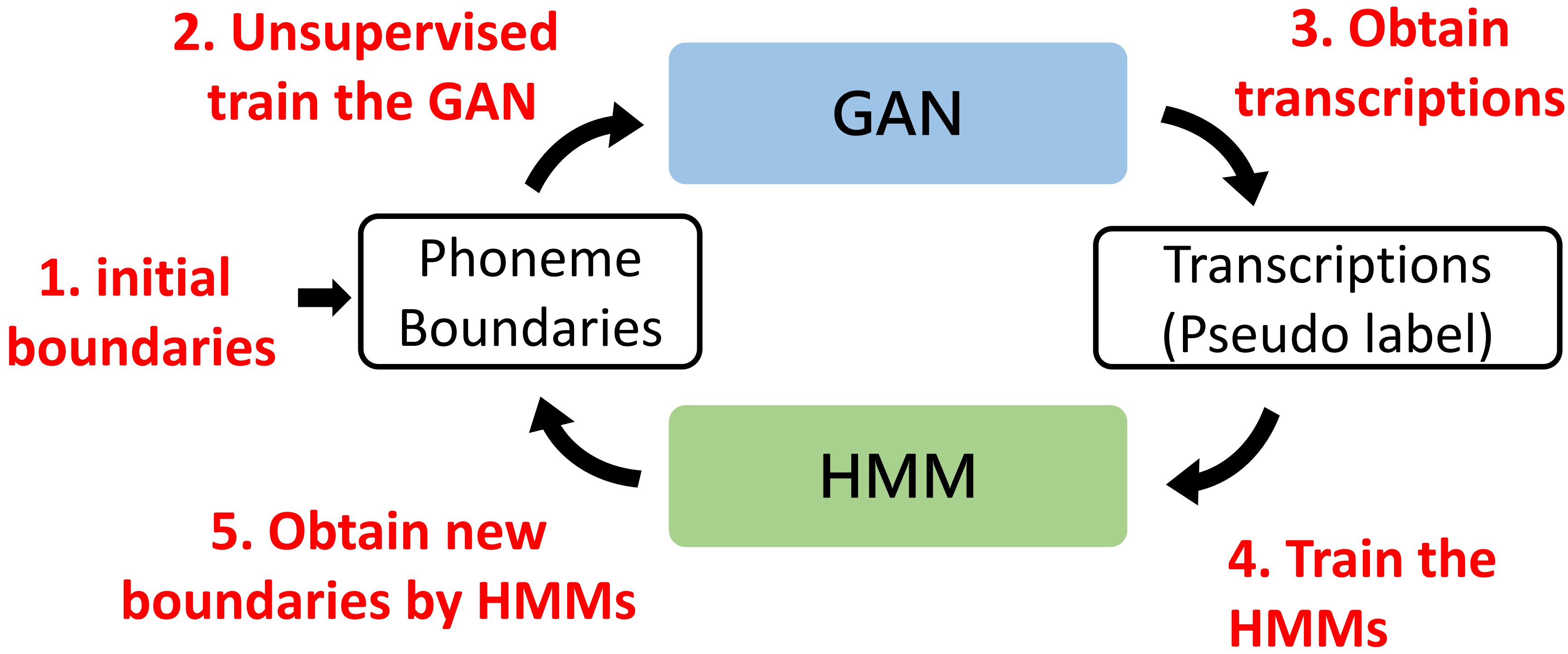


Inspired From [Yeh, et al., ICLR 2019]

Experimental Results [Chen, et al., INTERSPEECH, 2019]

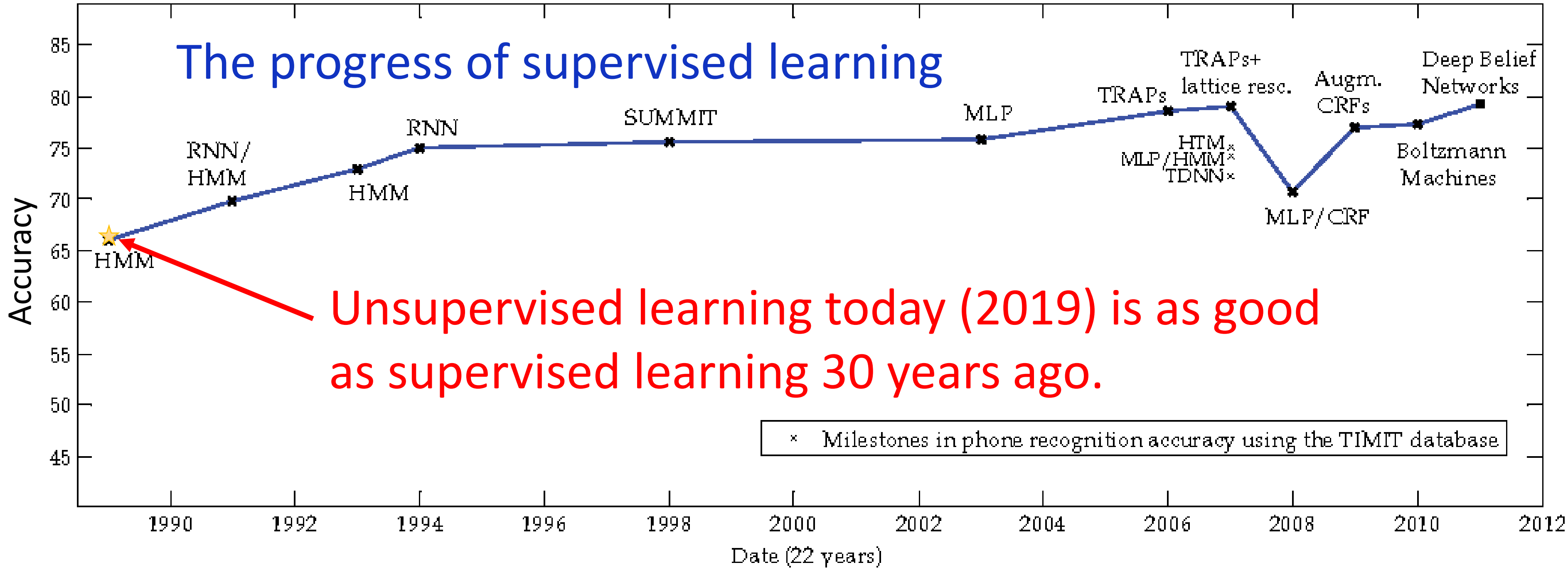
Approaches			PER	
			Matched	Nonmatched
Supervised				
RNN Transducer			17.7	-
Standard HMMs			21.5	-
Completely unsupervised (no label at all)				
Generator (v1)			76.0	-
Generator (v2)	Iteration 1	GAN	48.6	50.0

Refining Boundaries



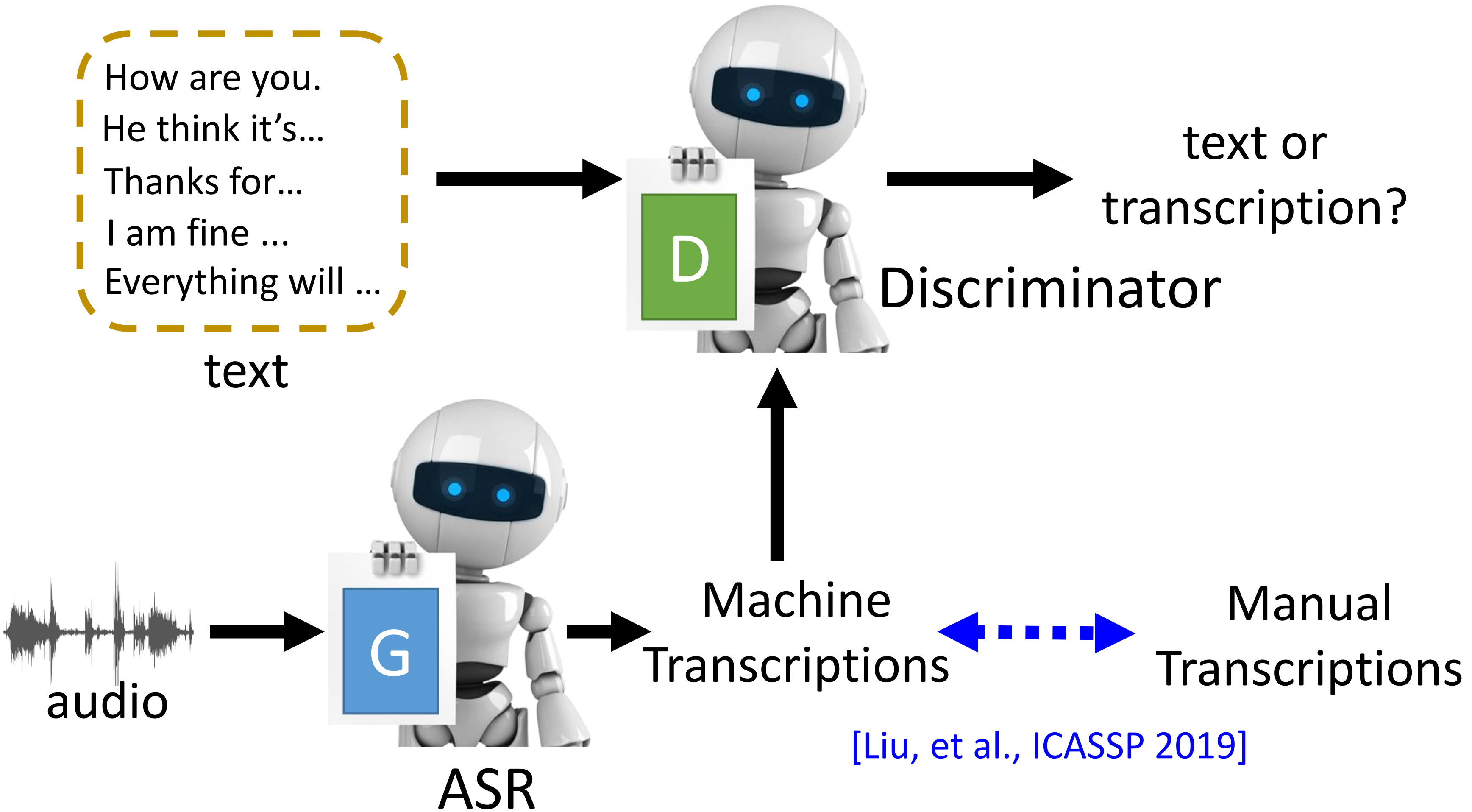
Experimental Results [Chen, et al., INTERSPEECH, 2019]

Approaches			PER	
			Matched	Nonmatched
Supervised				
RNN Transducer			17.7	-
Standard HMMs			21.5	-
Completely unsupervised (no label at all)				
Generator (v1)			76.0	-
Generator (v2)	Iteration 1	GAN	48.6	50.0
		HMM	30.7	39.5
	Iteration 2	GAN	41.0	44.3
		HMM	27.0	35.5
	Iteration 3	GAN	38.4	44.2
		HMM	26.1	33.1



The image is modified from: Phone recognition on the TIMIT database Lopes, C. and Perdigão, F., 2011. Speech Technologies, Vol 1, pp. 285--302.

Semi-supervised Speech Recognition

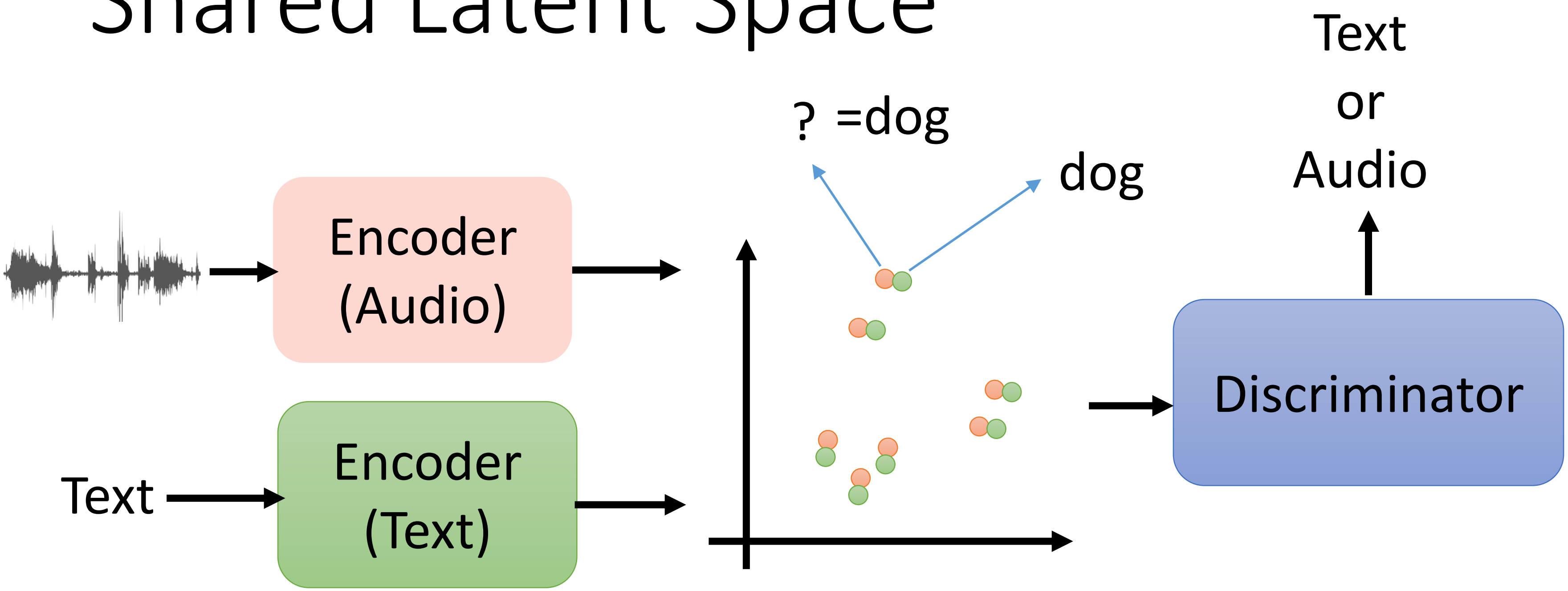


[Liu, et al., ICASSP 2019]

Using ***100 hours pairs annotated audio*** from Librispeech, and ***text without audio***

21.7% WER → 18.7% WER

Shared Latent Space



- Initial attempt [Chen, et al., SLT, 2018]
- 76.3% WER on Librispeech [Chung, et al., NIPS 2018]
- WSJ with 2.5 hours paired data: 64.6% WER
  [Jennifer Drexler, et al., SLT 2018]
- LJ speech with 20 mins paired data: 11.7% PER [Ren, et al., ICML 2019]
- Unsupervised speech translation is possible [Chung, et al., ICASSP 2019]

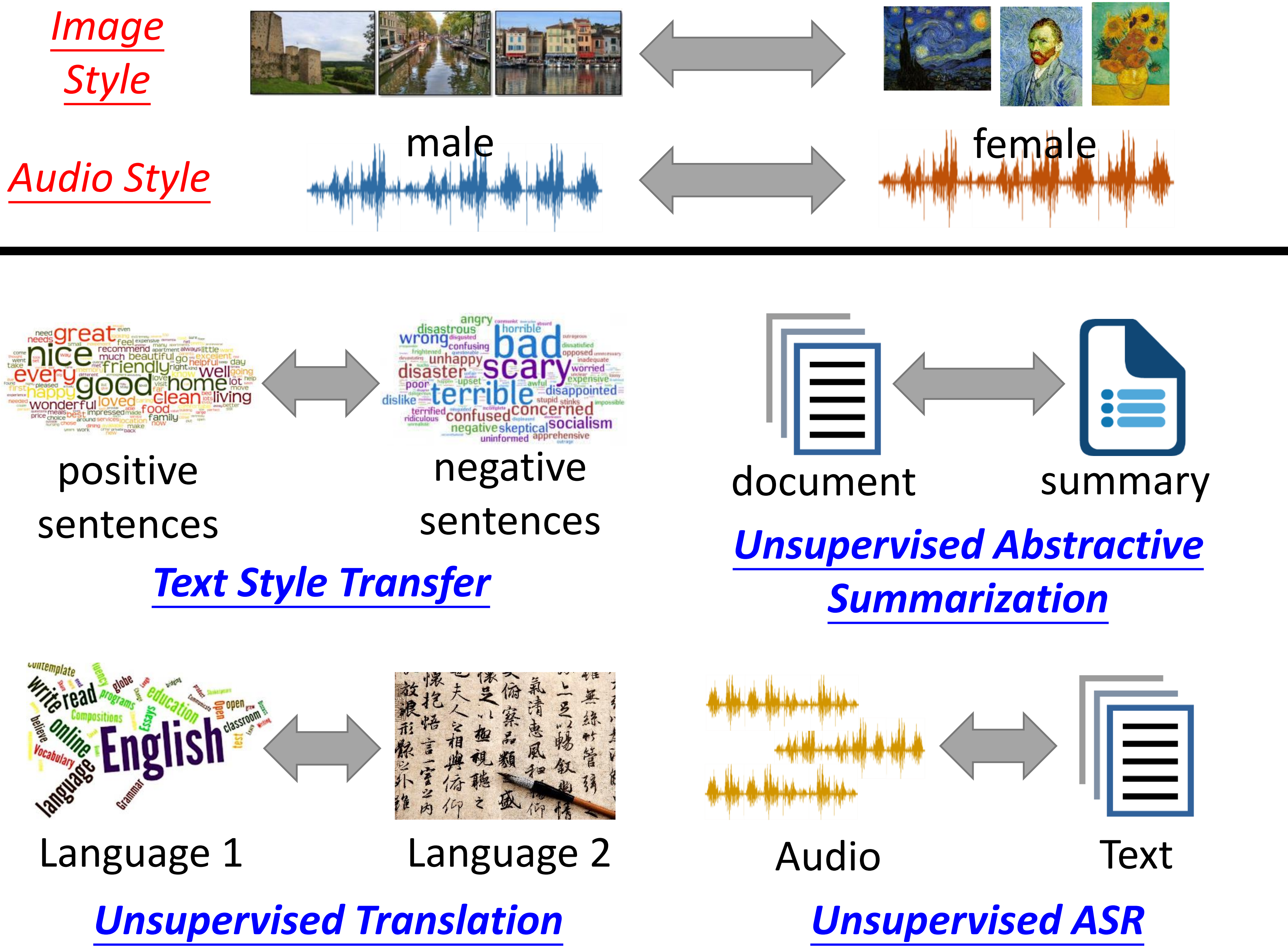


Image Style
male
female
Audio Style
positive sentences
negative sentences
Text Style Transfer
document
summary
Unsupervised Abstractive Summarization
Language 1
Language 2
Unsupervised Translation
Audio
Text
Unsupervised ASR

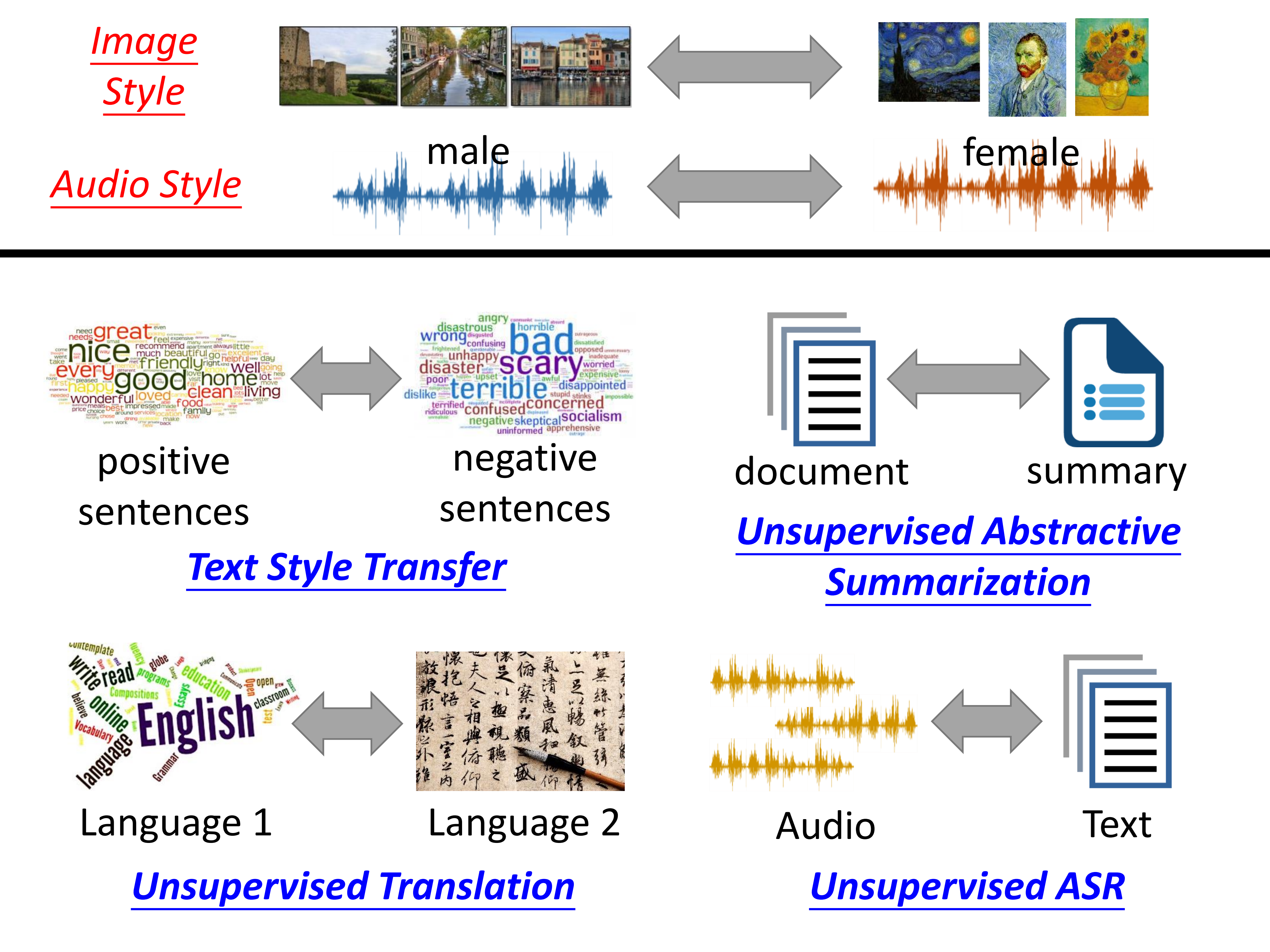


Image Style
male
female
Audio Style
positive sentences
negative sentences
Text Style Transfer
document
summary
Unsupervised Abstractive Summarization
Language 1
Language 2
Unsupervised Translation
Audio
Text
Unsupervised ASR

Reference

- [Lee, et al., ICASSP’18] Chih-Wei Lee, Yau-Shian Wang, Tsung-Yuan Hsu, Kuan-Yu Chen, Hung-Yi Lee, Lin-shan Lee, Scalable Sentiment for Sequence-to-sequence Chatbot Response with Performance Analysis, ICASSP, 2018
- [Ning Dai, et al., ACL’19] Ning Dai, Jianze Liang, Xipeng Qiu, Xuanjing Huang, Style Transformer: Unpaired Text Style Transfer without Disentangled Latent Representation, ACL, 2019
- [Lample, et al., ICLR’19] Guillaume Lample, Sandeep Subramanian, Eric Smith, Ludovic Denoyer, Marc'Aurelio Ranzato, Y-Lan Boureau, Multiple-Attribute Text Rewriting, ICLR, 2019
- [Lample, et al., EMNLP’18] Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, Marc’Aurelio Ranzato, Phrase-Based & Neural Unsupervised Machine Translation, EMNLP,2018

Reference

- [Hu, et al., ICML’17] Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, Eric P. Xing, Toward Controlled Generation of Text, ICML, 2017
- [Fu, et al., AAAI’17] Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan, “Style transfer in text: Exploration and evaluation,” AAAI, 2017.
- [Shen, et al., NIPS’17] Tianxiao Shen, Tao Lei, Regina Barzilay, Tommi Jaakkola, Style Transfer from Non-Parallel Text by Cross-Alignment, NIPS, 2017
- [Li, et al., NAACL’18] Juncen Li, Robin Jia, He He, Percy Liang, Delete, Retrieve, Generate: a Simple Approach to Sentiment and Style Transfer, NAACL, 2018
- [Xu, et al., ACL’18] Jingjing Xu, Xu Sun, Qi Zeng, Xuancheng Ren, Xiaodong Zhang, Houfeng Wang, Wenjie Li, Unpaired Sentiment-to-Sentiment Translation: A Cycled Reinforcement Learning Approach, ACL, 2018

Reference

- [Wang, Lee, EMNLP'18] Yau-Shian Wang, Hung-Yi Lee, "Learning to Encode Text as Human-Readable Summaries using Generative Adversarial Networks", EMNLP, 2018
- [Chu, et al., ICML'19] Eric Chu, Peter Liu, "MeanSum: A Neural Model for Unsupervised Multi-Document Abstractive Summarization", ICML, 2019
- [Baziotis, et al., NAACL'19] Christos Baziotis, Ion Androutsopoulos, Ioannis Konstas, Alexandros Potamianos, "SEQ^3: Differentiable Sequence-to-Sequence-to-Sequence Autoencoder for Unsupervised Abstractive Sentence Compression", NAACL 2019

Reference

- [Conneau, et al., ICLR’18] Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, Hervé Jégou, Word Translation Without Parallel Data, ICLR 2018
- [Lample, et al., ICLR’18] Guillaume Lample, Ludovic Denoyer, Marc'Aurelio Ranzato, Unsupervised Machine Translation Using Monolingual Corpora Only, ICLR, 2018
- [Lample, et al., NeurIPS’19] Guillaume Lample, Alexis Conneau, Cross-lingual Language Model Pretraining, NeurIPS, 2019
- [Artetxe, et al., ACL’18] Mikel Artetxe, Gorka Labaka, and Eneko Agirre, A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings, ACL, 2018
- [Hartmann, et al., NeurIPS’19] Mareike Hartmann, Yova Kementchedjhieva, Anders Søgaard, Comparing Unsupervised Word Translation Methods Step by Step, NeurIPS, 2019

Reference

- [Li, et al., EMNLP, 2017] Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter, Dan Jurafsky, Adversarial Learning for Neural Dialogue Generation, EMNLP, 2017
- [Matt J. Kusner, et al., arXiv, 2016] Matt J. Kusner, José Miguel Hernández-Lobato, GANS for Sequences of Discrete Elements with the Gumbel-softmax Distribution, arXiv 2016
- [Tong Che, et al, arXiv, 2017] Tong Che, Yanran Li, Ruixiang Zhang, R Devon Hjelm, Wenjie Li, Yangqiu Song, Yoshua Bengio, Maximum-Likelihood Augmented Discrete Generative Adversarial Networks, arXiv 2017
- [Yu, et al., AAAI, 2017] Lantao Yu, Weinan Zhang, Jun Wang, Yong Yu, SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient, AAAI 2017
- [Sai Rajeswar, et al., arXiv, 2017] Sai Rajeswar, Sandeep Subramanian, Francis Dutil, Christopher Pal, Aaron Courville, Adversarial Generation of Natural Language, arXiv, 2017

Reference

- [Zhen Xu, et al., EMNLP, 2017] Zhen Xu, Bingquan Liu, Baoxun Wang, Chengjie Sun, Xiaolong Wang, Zhuoran Wang, Chao Qi , Neural Response Generation via GAN with an Approximate Embedding Layer, EMNLP, 2017
- [Alex Lamb, et al., NIPS, 2016] Alex Lamb, Anirudh Goyal, Ying Zhang, Saizheng Zhang, Aaron Courville, Yoshua Bengio, Professor Forcing: A New Algorithm for Training Recurrent Networks, NIPS, 2016
- [Yizhe Zhang, et al., ICML, 2017] Yizhe Zhang, Zhe Gan, Kai Fan, Zhi Chen, Ricardo Henao, Dinghan Shen, Lawrence Carin, Adversarial Feature Matching for Text Generation, ICML, 2017
- [Jiaxian Guo, et al., AAAI, 2018] Jiaxian Guo, Sidi Lu, Han Cai, Weinan Zhang, Yong Yu, Jun Wang, Long Text Generation via Adversarial Training with Leaked Information, AAAI, 2018
- [Kevin Lin, et al, NIPS, 2017] Kevin Lin, Dianqi Li, Xiaodong He, Zhengyou Zhang, Ming-Ting Sun, Adversarial Ranking for Language Generation, NIPS, 2017

Reference

- [Tual, Lee, TASLP, 2019]Yi-Lin Tuan, Hung-Yi Lee, Improving Conditional Sequence Generative Adversarial Networks by Stepwise Evaluation, TASLP, 2019
- [Xu, et al., EMNLP, 2018] Jingjing Xu, Xuancheng Ren, Junyang Lin, Xu Sun, Diversity-Promoting GAN: A Cross-Entropy Based Generative Adversarial Network for Diversified Text Generation, EMNLP, 2018
- [Weili Nie, et al. ICLR, 2019] Weili Nie, Nina Narodytska, Ankit Patel, RelGAN: Relational Generative Adversarial Networks for Text Generation, ICLR 2019
- [Ofir Press, et al., ICML workshop, 2017]Ofir Press, Amir Bar, Ben Bogin, Jonathan Berant, Lior Wolf, Language Generation with Recurrent Generative Adversarial Networks without Pre-training, ICML workshop, 2017
- [William Fedus, et al., ICLR, 2018] William Fedus, Ian Goodfellow, Andrew M. Dai, MaskGAN: Better Text Generation via Filling in the______, ICLR, 2018
- [Cyprien de Masson d'Autume, et al., arXiv 2019] Cyprien de Masson d'Autume, Mihaela Rosca, Jack Rae, Shakir Mohamed, Training language GANs from Scratch, arXiv 2019

Reference

- [Liu, et al., ICASSP’18] Alexander H. Liu, Hung-yi Lee, Lin-shan Lee, Adversarial Training of End-to-end Speech Recognition Using a Criticizing Language Model, ICASSP 2018
- [Liu, et al., INTERSPEECH’18] Da-Rong Liu, Kuan-Yu Chen, Hung-Yi Lee, Lin-shan Lee, Completely Unsupervised Phoneme Recognition by Adversarially Learning Mapping Relationships from Audio Embeddings, INTERSPEECH, 2018
- [Chen, et al., INTERSPEECH’19] Kuan-yu Chen, Che-ping Tsai, Da-Rong Liu, Hung-yi Lee and Lin-shan Lee, "Completely Unsupervised Phoneme Recognition By A Generative Adversarial Network Harmonized With Iteratively Refined Hidden Markov Models", INTERSPEECH, 2019
- [Chen, et al., SLT’18] Yi-Chen Chen, Sung-Feng Huang, Chia-Hao Shen, Hung-yi Lee, Lin-shan Lee, "Phonetic-and-Semantic Embedding of Spoken Words with Applications in Spoken Content Retrieval", SLT, 2018
- [Yeh, et al., ICLR’19] Chih-Kuan Yeh, Jianshu Chen, Chengzhu Yu, Dong Yu, Unsupervised Speech Recognition via Segmental Empirical Output Distribution Matching, ICLR, 2019

Reference

- [Chung, et al., NIPS 2018] Yu-An Chung, Wei-Hung Weng, Schrasing Tong, James Glass, Unsupervised Cross-Modal Alignment of Speech and Text Embedding Spaces, NIPS, 2018
- [Chung, et al., ICASSP 2019] Yu-An Chung, Wei-Hung Weng, Schrasing Tong, James Glass, Towards Unsupervised Speech-to-Text Translation, ICASSP 2019
- [Ren, et al., ICML 2019]Yi Ren, Xu Tan, Tao Qin, Sheng Zhao, Zhou Zhao, Tie-Yan Liu, Almost Unsupervised Text to Speech and Automatic Speech Recognition, ICML 2019
- [Jennifer Drexler, et al., SLT 2018] Jennifer Drexler, James R. Glass, “Combining End-to-End and Adversarial Training for Low-Resource Speech Recognition”, SLT 2018