Digital Data Flows Masterclass on **Machine Learning and Speech**

9 December 2020

FUTURE OF PRIVACY FORUM

Future of Privacy Forum

The Supporters

150+	25+	15+	5
Companies	Leading Academics	Advocates and Civil Society	Foundations

The Mission

Bridging the policymaker-industry-academic gap in privacy policy

Developing privacy protections, ethical norms, & responsible business practices

The Workstreams

AI & Ethics
Student Data

Apps & Ad Tech
Mobility & Location

Privacy Enhancing Tech
Smart Communities

DIGITAL DATA FLOWS

EMERGING TECH MASTERCLASS SERIES

The **Digital Data Flows Masterclass** series is an educational program designed for regulators, policymakers, and staff seeking to better understand the data-driven technologies at the forefront of privacy and data protection law & policy.

Classes	Date	
1. Artificial Intelligence and Machine Learning	25 Oct 2018	Brussels
2. Location Data: GPS, Wi-Fi, &andSpatial Analytics	27 Nov 2018	Brussels
3. De-Identification, Differential Privacy, and Homomorphic Encryption	30 Jan 2019	Brussels
4. Online Advertising, Data Flows, Behavioral Targeting, and Cross-Device Tracking	1 May 2019	Wash. DC
5. Mobile Apps: Operating Systems, Software Development Kits (SDKs), and User Controls	25 Jul 2019	Virtual
6. Facial Recognition and Biometric Data	27 Feb 2020	Wash. DC
7. Connected Cars and Autonomous Vehicles	25 Jun 2020	Virtual
8. Blockchain Technoloties	29 Oct 2020	Virtual
9. Machine Learning and Speech	9 Dec 2020	Virtual

Access recordings and materials for all previous classes at www.fpf.org/classes

Speakers

Professor Marine Carpuat, Associate Professor in Computer Science at the University of Maryland

Dr. Prem Natarajan, VP, Alexa AI-NU

Neural Machine Translation for Communication Across Language Barriers

Marine Carpuat

marine@cs.umd.edu

Future of Privacy Forum

Digital Data Flows Masterclass 2020/12/09

7,111 languages spoken today

23 account for 50% of population

43% endangered

Source: Ethnologue

UMD Machine Translation
@WMT 2018

这座中国首都拥有速度高得惊人**的互**联网，有人脸识别软件等尖**端技**术，在人工智能方面投入了**巨**资并且拥有无可匹敌的国际化**能量，** **它**对富于探索精神的外国**人而言是最激**动人心的城市之一。

The Chinese capital, with its surprisingly high-speed Internet, sophisticated technology such as face-recognition software, has invested heavily in artificial intelligence and has unrivaled international energy, and is one of the most exciting cities for exploration-minded foreigners.

[Xu & Carpuat WMT 2018]

Translation as Machine Learning

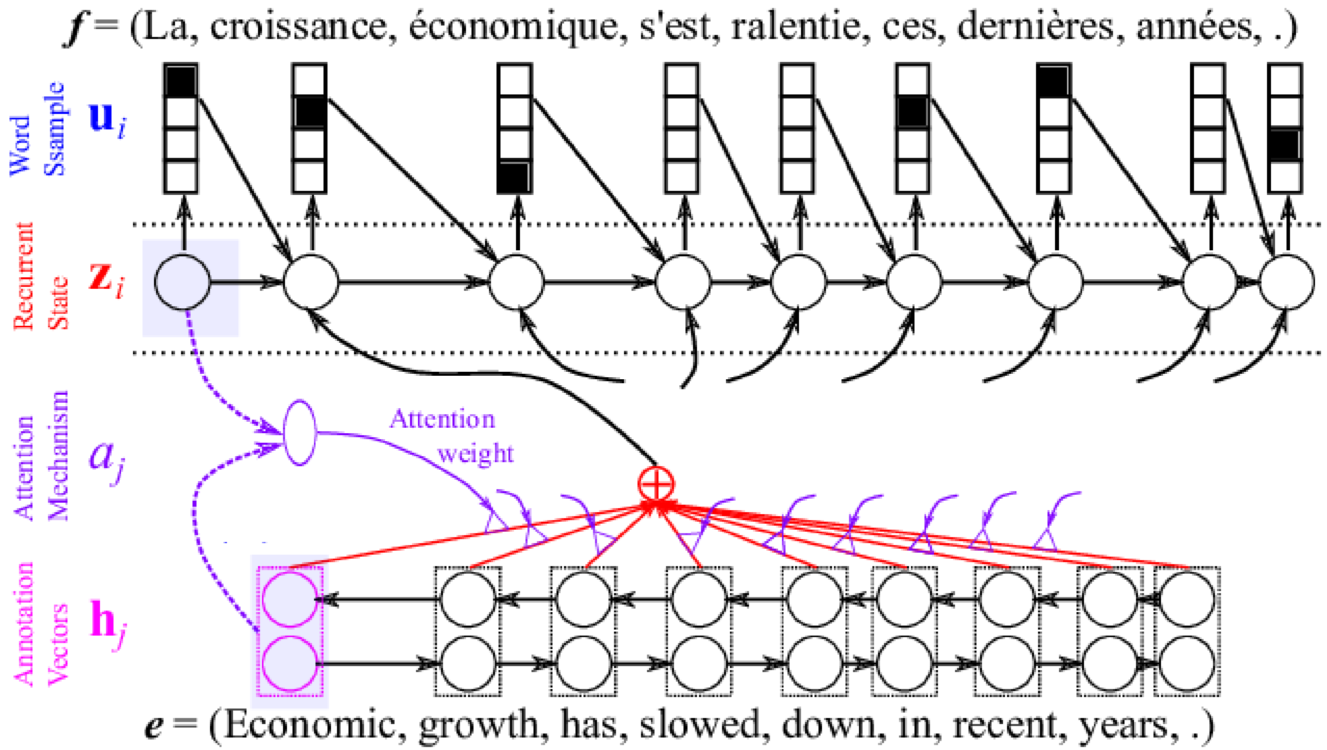
An English sentence e is translated into the French sentence

$$f^* = \text{argmax}_f \, p(f|e;\theta)$$

"Learning" refers to finding good values for the model parameters

$$\theta^* = \text{argmax}_\theta \sum_i \log p(f_i \, |e_i;\theta)$$

Translation as Deep Learning



$$p(f \mid e;\ \theta) = \prod_{t=1}^{T} p(f_t \mid f_{<t}, e;\ \theta)$$

Image: Kyunghung Cho

Challenges of Translation as Deep Learning

requires millions of translation examples

not available for many languages!

raises fundamental machine learning challenges

intractably large output space, infinitely many correct outputs…

makes errors that have real world impact

yet models are opaque, and developed independency from use cases

Beyond Translation

sequence-to-sequence models are general purpose tools to generate language

used for dialog, text summarization, question answering, style transfer, etc.

share architecture design with language models and word representation models (aka word embeddings)

How can we train neural models with limited translation data?

Learning Paradigms to Exploit Diverse Data Sources

Supervised learning from parallel samples (translations)

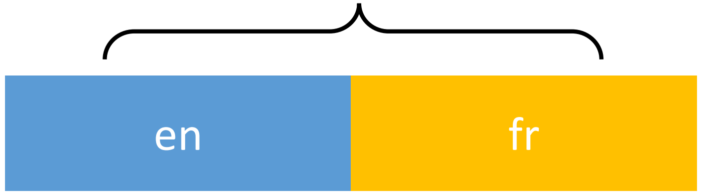
TED

Unsupervised learning from unpaired monolingual samples

Supervised Learning

requires parallel samples: input sequence paired with a correct output sequence

$$\theta^* = \operatorname{argmax}_{\theta} \sum_{i} \log p(f_i \mid e_i; \theta)$$



Unsupervised Learning

Learn a **language model** per language which predicts the next word given previous words in a sentence

fr

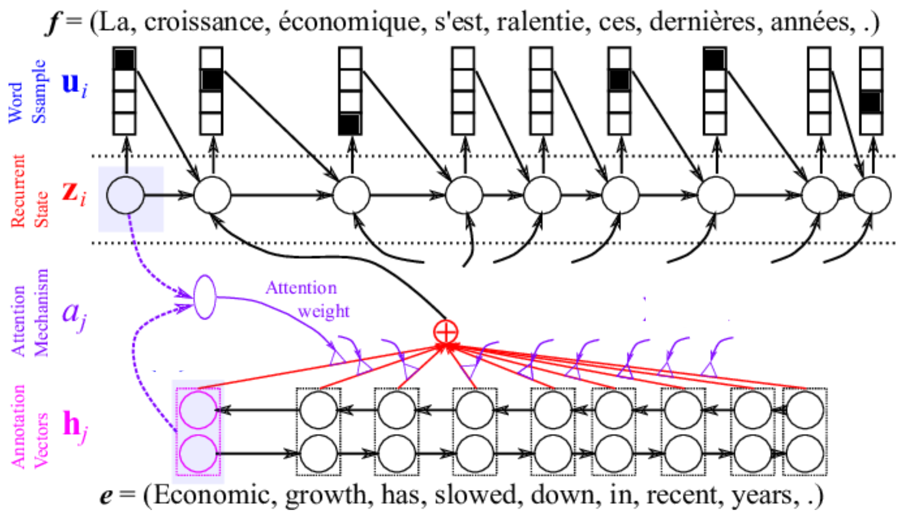
$$\theta_f{}^* = \text{argmax}_\theta \sum_f \log p(f_i \mid f_{<i}; \theta)$$

en

$$\theta_e{}^* = \text{argmax}_\theta \sum_e \log p(e_i \mid e_{<i}; \theta)$$

Unsupervised Learning

Learn a **language model** per language which predicts the next word given previous words in a sentence



+ dictionaries to capture cross-lingual correspondences

Semi-Supervised Learning

learn from (few) paired + (many) unpaired samples

common approach: **back translation**

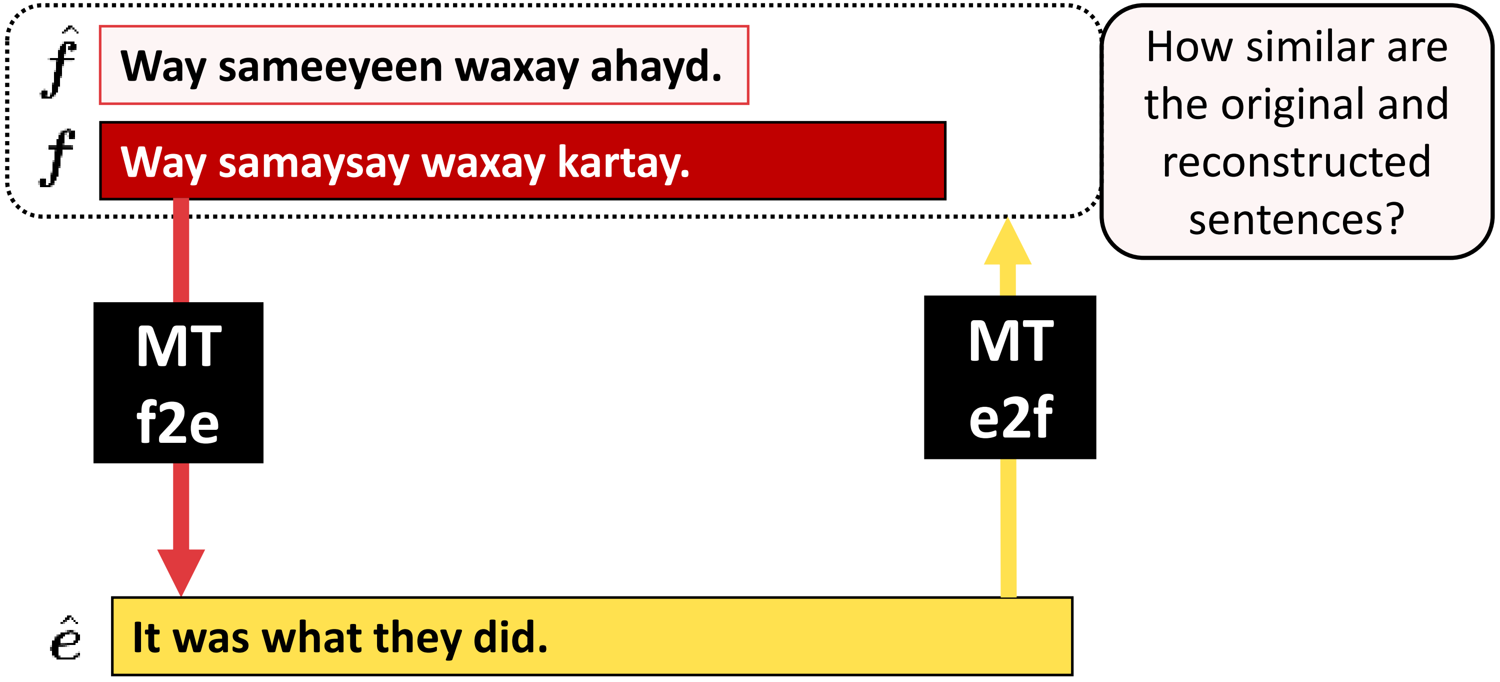
use auxiliary Machine Translation (MT) to translate target into source

en	fr
en* = MT(fr)	fr

simple and effective [Sennrich et al. 2016]

Semi-Supervised Learning

Alternative: **reconstruction**



More complex than backtranslation,
but see Niu et al. [2019] for efficient variants

Provides a theoretical framework to analyze semi-supervised learning [Xu et al., 2020]

(Semi-) supervised learning is key to recent MT successes


THE VERGE
GOOGLE WEB APPS
Google's AI translation system is approaching human-level accuracy

ZDNet
VIDEOS 5G WINDOWS 10 CLOUD INNOVATION SECURITY TECH PRO MORE NEWSLETTERS
Microsoft researchers match human levels in translating news from Chinese to English
Microsoft researchers are making progress in translating text using deep neural networks AI training techniques.
By Mary Jo Foley for All About Microsoft | March 14, 2018 -- 15:06 GMT (08:06 PDT) | Topic: Artificial Intelligence

2016

2018

Better algorithms are still needed to scale to all languages

How can we make neural models more responsive to user needs?

Neural MT still fails, sometimes catastrophically

HAARETZ | Israel News | All sections | Israel - BDS | Israel settlements | Italy - anti-Semitism | Flak

Home > Israel News

Israel Arrests Palestinian Because Facebook Translated 'Good Morning' to 'Attack Them'

No Arabic-speaking police officer read the post before arresting the man, who works at a construction site in a West Bank settlement

Yotam Berger | Oct 22, 2017 1:36 PM

https://www.haaretz.com/israel-news/palestinian-arrested-over-mistranslated-good-morning-facebook-post-1.5459427

Fluently inadequate translations

One type of catastrophic failure

An **adequate** translation accurately conveys the meaning of the source

A **fluent** translation is well formed in the output language

[Martindale et al. MT Summit 2019]

Fluently inadequate translations

Kanlar içindeydi	**Fluent**	**Disfluent**
Adequate	He was covered in blood.	There was blood on.
Inadequate	Drinking blood.	Inthe Kanlar.

Actual system outputs from WMT16 News TR-EN translation task. From left-to-right and top-to-bottom: online-G, online-B, dvorkanton, jhu

Fluently inadequate translations

Can we detect them automatically?

How often do they occur?

For what types of systems are they most frequent?

Quantifying Fluency

We score fluency of a sentence using its language model probability

$$p(e\ ;\theta) = \prod_i p(e_i\,|e_{<i};\theta)$$

(modified to account for sentence length + outliers)

	CS-EN	DE-EN	FI-EN	RO-EN	RU-EN	TR-EN	All
Percent fluent	59.22%	59.70%	56.79%	58.04%	60.80%	48.81%	57.21%
Precision	65.35	63.36	59.62	62.06	66.22	52.37	61.42
Recall	90.97	87.29	91.56	92.20	87.67	87.77	89.38
F1	76.06	73.42	72.21	74.18	75.45	65.60	72.81

Classifying fluent vs. disfluent sentences

Quantifying Adequacy

Approach: compute semantic similarity between MT and human translation

"BLEU": counts n-gram matches

"BVSS": combines pairwise word similarity into sentence score

	Prec.	**Recall**	**F1**
BLEU	94.33	99.08	96.65
Averaged Embeddings	84.56	**99.15**	91.28
BVSS	**99.39**	99.04	**99.22**
BVSS-Reference	99.00	99.03	99.01
BVSS-System	99.17	99.03	99.10
BLEU+BVSS	**99.61**	**99.81**	**99.71**

Classifying adequate vs. inadequate translations

System Level Analysis

we study 36 MT systems spanning

3 training configurations

General (large, out-of-domain), TED (small, in-domain), Adapted (general model adapted to TED)

2 paradigms

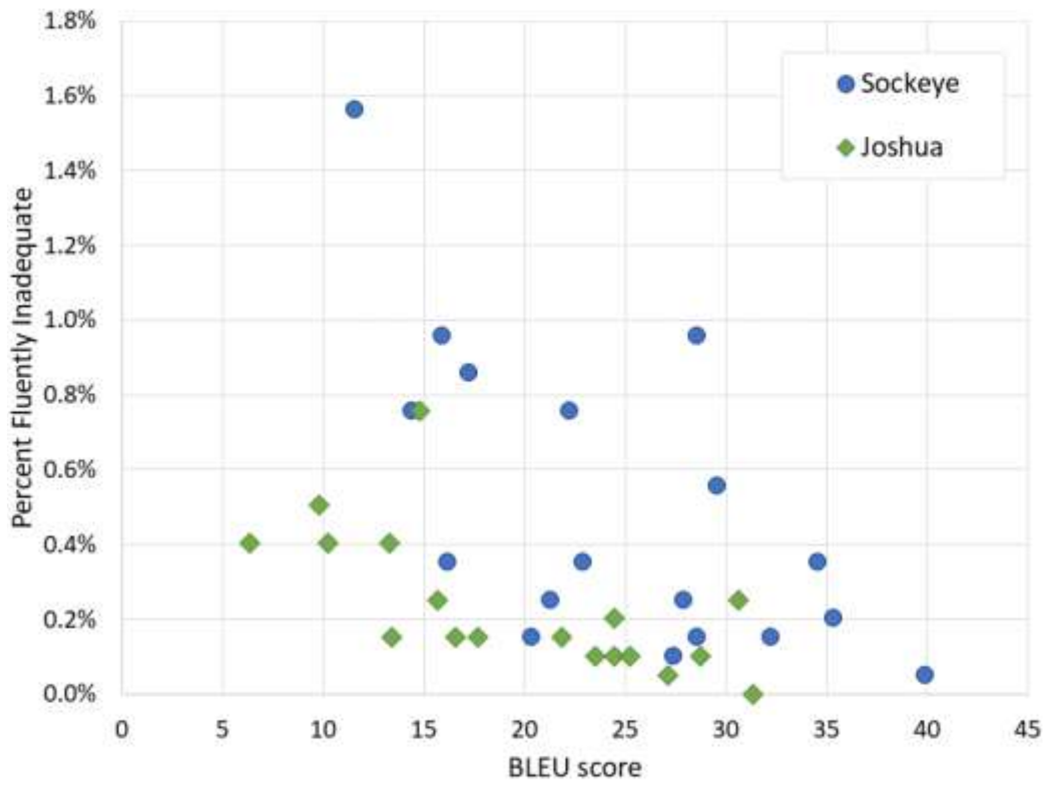
Statistical MT (Joshua), Neural MT (Sockeye)

6 language pairs

	Arabic	**Chinese**	**Farsi**	**German**	**Korean**	**Russian**
Joshua General	23.50	30.65	13.41	6.34	24.49	14.79
Joshua TED Only	24.49	28.72	16.56	9.81	21.85	13.32
Joshua Adapted	27.11	31.35	17.71	10.24	25.23	15.70
Sockeye General	29.6	34.59	22.22	11.56	**28.6**	15.92
Sockeye TED Only	27.42	32.25	21.31	14.4	22.9	16.18
Sockeye Adapted	**35.37**	**39.9**	**27.92**	**17.22**	**28.6**	**20.37**

Table 6: BLEU scores for all systems

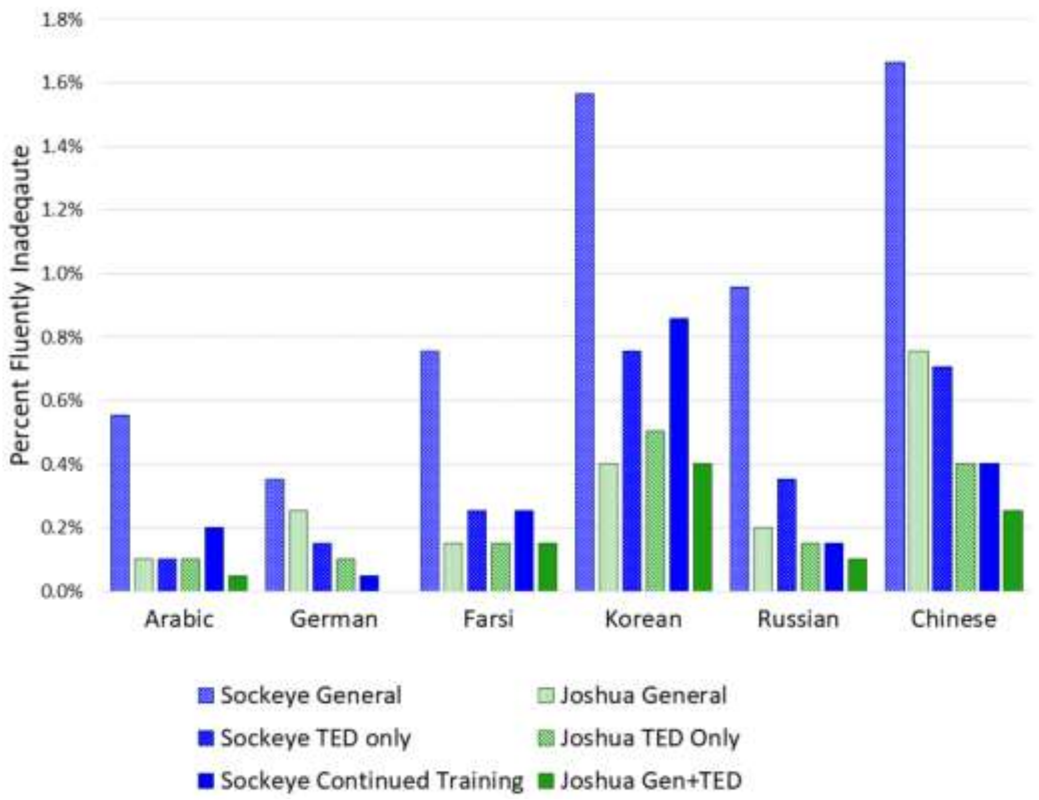
Fluently inadequate translations



For the same quality, neural models have more fluently inadequate than older statistical models

Percent fluently inadequate for neural models improves rapidly as BLEU improves

Fluently inadequate translations



Neural models have consistently higher percent fluently inadequate than statistical models

Domain adaptation improves score for both, but more dramatically for neural models

Even highest percent fluently inadequate is <2%

Fluently inadequate translations

are infrequent but important

more likely to mislead users

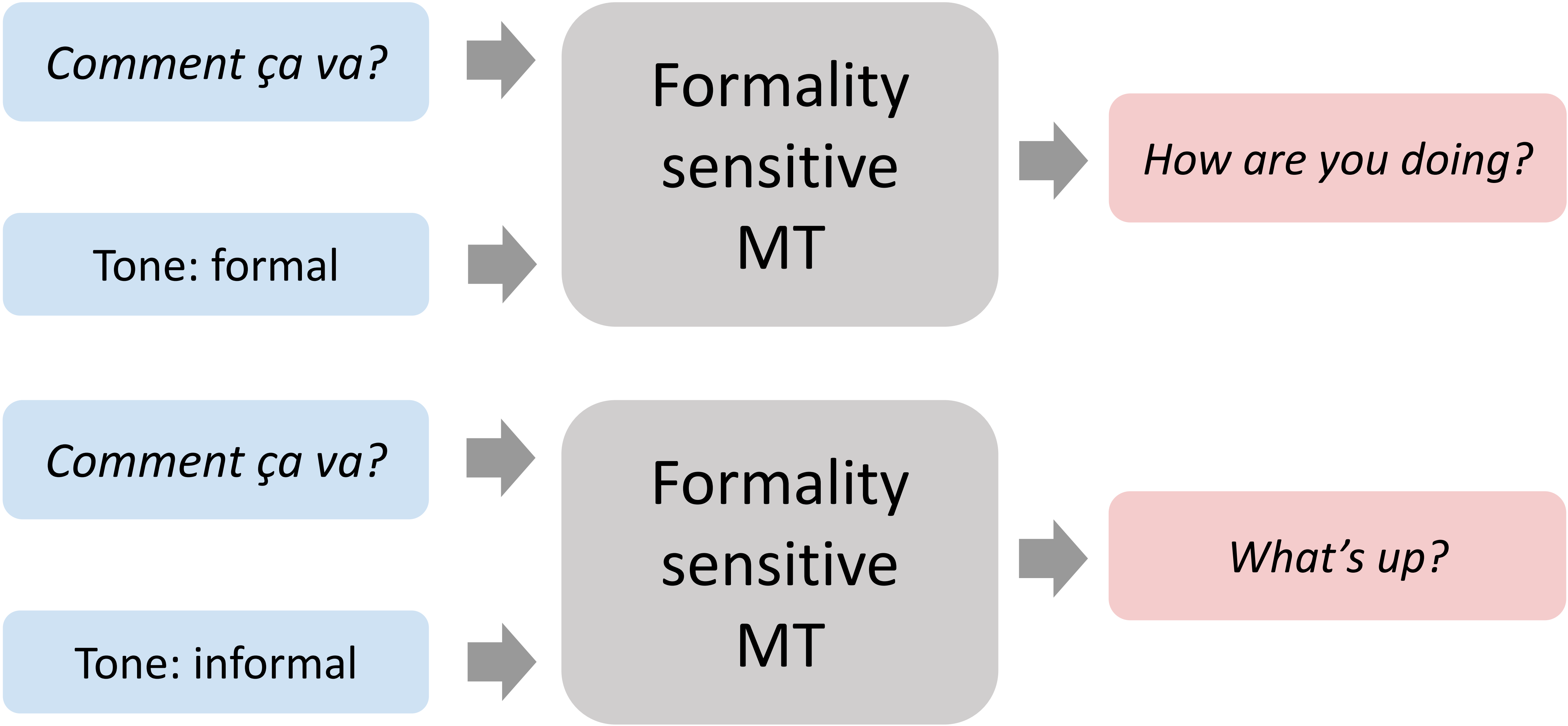
millions of translation requests per day!

affect all models considered

happen more often for neural models, when translating out of domain, and low-resource languages

How can we make neural models more responsive to user needs?

Can We Make MT Sensitive to Formality?



[Niu & Carpuat, EMNLP 2017]

Can We Control the Complexity Of MT?

El museo Mauritshuis abre una exposición dedicada a los autorretratos del siglo XVII.

Audience: fluent English speaker

Complexity Controlled MT

The Mauritshuis museum is staging an exhibition focused solely on 17th century self-portraits.

El museo Mauritshuis abre una exposición dedicada a los autorretratos del siglo XVII.

Audience: 2nd language learner

Complexity Controlled MT

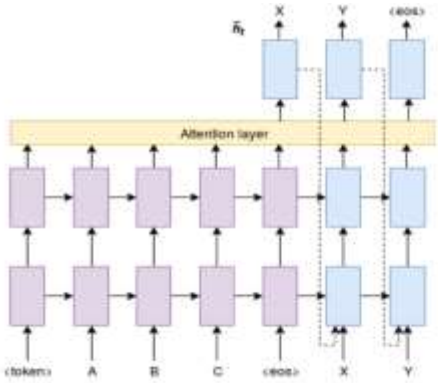
The Mauritshuis museum is going to show self-portraits.

[Agrawal & Carpuat, EMNLP 2019]

Adapting translation output to different audiences via multi-task learning

English or Spanish input

Desired style
e.g. reading grade level



English output

$$\text{Multi-task loss} = \sum_{(s_i,g_s,e_o} \underbrace{\log P(e_o|s_i,g_e;\theta)}_{L_{CMT}} + \sum_{(e_i,g_e,e_o} \underbrace{\log P(e_o|e_i,g_e;\theta)}_{L_{Simplify}} + \sum_{(s_i,e_o} \underbrace{\log P(e_o|s_i;\theta)}_{L_{MT}}$$

L_{CMT}: Spanish sentences translated into simpler English

$L_{Simplify}$: Complex English sentences paired with simpler English

L_{MT}: Spanish-English translation examples

Tailoring MT output to different audiences

Possible with multi-task models

Outperform pipeline models

Powerful framework to incorporate diverse data sources

But open questions remain

How to better control the nature of style transfer operations? Can we relax data requirements with unsupervised methods?

Deep neural networks provide a powerful framework to model translation

Semi-supervised learning is key to train high-performing models

How can we make neural models more responsive to user needs?

Neural Machine Translation for Communication Across Language Barriers

Marine Carpuat

marine@cs.umd.edu

Future of Privacy Forum

Digital Data Flows Masterclass 2020/12/09

References

Weijia Xu, Xing Niu and Marine Carpuat. "Dual Reconstruction: a Unifying Objective for Semi-Supervised Neural Machine Translation". Findings of EMNLP 2020.

Sweta Agrawal and Marine Carpuat. "Controlling Text Complexity in Neural Machine Translation". EMNLP 2019.

Marianna J. Martindale, Marine Carpuat, Kevin Duh and Paul McNamee. "Identifying Fluently Inadequate Output in Neural and Statistical Machine Translation". Machine Translation Summit. 2019.

Xing Niu, Weijia Xu and Marine Carpuat. "Bi-Directional Differentiable Input Reconstruction for Low-Resource Neural Machine Translation". ACL 2019.

Xing Niu, Sudha Rao and Marine Carpuat. "Multi-Task Neural Models for Translating Between Styles Within and Across Languages". COLING 2018.

Weijia Xu and Marine Carpuat. "The University of Maryland's Chinese-English Neural Machine Translation Systems at WMT18". WMT 2018.

Marianna J. Martindale and Marine Carpuat. "Fluency Over Adequacy: A Pilot Study in Measuring User Trust in Imperfect MT". AMTA 2018.

Rico Sennrich, Barry Haddow, Alexandra Birch. "Improving Neural Machine Translation Models with Monolingual Data". ACL 2016

Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio. "Neural Machine Translation by Jointly Learning to Align and Translate". ICLR 2015.

Thank you!

Contact us:
info@fpf.org
@FutureofPrivacy
www.fpf.org/classes

FUTURE OF PRIVACY FORUM