

Lecturer Introduction



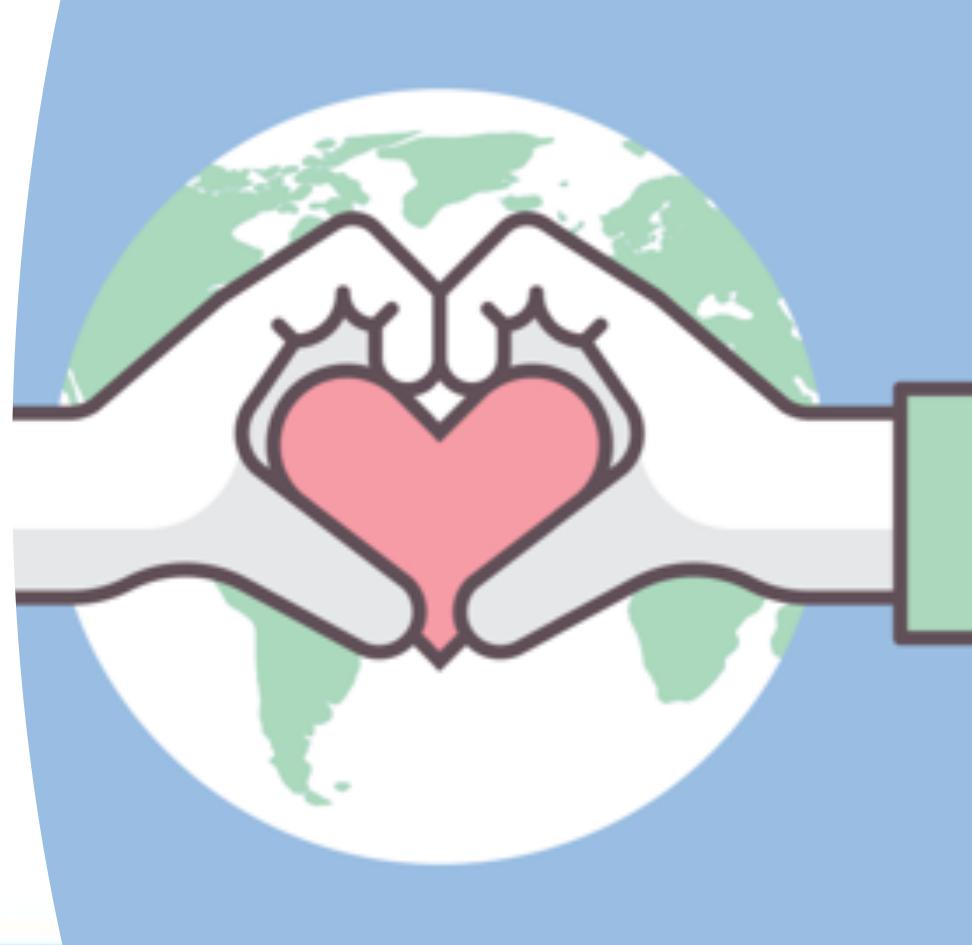
Francisco (Paco) Guzmán

Paco is a Research Scientist Manager at Facebook AI where he leads a team working on People-Facing Translations. His research has been focused on several aspects of Machine Translation including low-resource translation, translation mining, evaluation and quality estimation.

Before joining Facebook in 2016, Paco was a Research Scientist at Qatar Computing Research Institute in Qatar in 2012-2016. He obtained his PhD in 2011 from ITESM (Monterrey Tech) in Mexico.

Translation at Facebook

Translation connects people



Translation at Facebook

FACEBOOK AI

Georgia Tech

6B+ daily translations



Translation at Facebook

FACEBOOK AI

Georgia Tech

Native

Jole Simmons
33 mins · 1 person

Congratulations to the newlyweds! I love you!
#maxiejohnsonwedding2017

National Cherry Blossom Festival
2 hr · 1 person

今週末に来てください。花は満開ですし、絶対に素晴らしいです！

Like Share

Le Petit Chat
1 hr · 1 person

Regardez comment nous fabriquons nos délicieux croissants. #beurre!

Le secret est le beurre!

Share

Bill Russell, Joe Tony and 80 others

Like Comment Share

What an English Speaker Sees

Jole Simmons
33 mins · 1 person

Congratulations to the newlyweds! I love you!
#maxiejohnsonwedding2017

National Cherry Blossom Festival
2 hr · 1 person

Please come down and join us this weekend. The flowers are in full bloom and absolutely stunning!

Like Share

Le Petit Chat
1 hr · 1 person

Take a look how we make our delicious croissants. #butter!

The secret is butter!

Share

Bill Russell, Joe Tony and 80 others

Like Comment Share

What a Turkish Speaker Sees

Jole Simmons
33 dk. · 1 person

Yeni evilere tebrikler! Sizler seviyoruz! #maxiejohnsonwedding2017

National Cherry Blossom Festival
2 saat · 1 person

Lütfen aşağı gelin ve bu hafta sonu bize katılın. Çiçekler tamamen açtılar ve çok etkileyiciler!

Like Share

Le Petit Chat
1 saat · 1 person

Gelin nasıl lezzetli kruvasanlar yaptığımıza bakın. #tereyag!

İşin sırrı tereyağında!

Paylaş

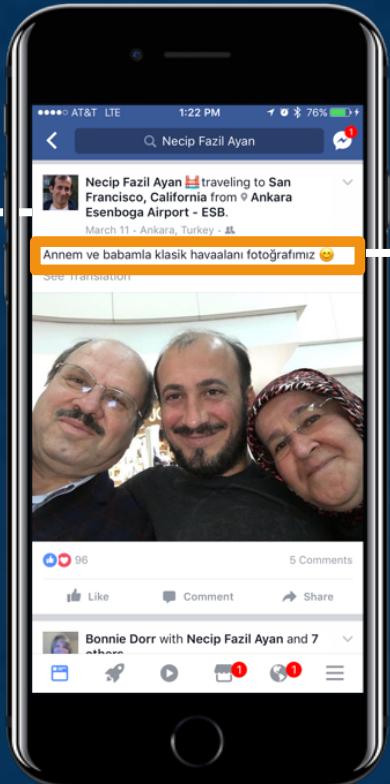
Bill Russell, Joe Tony ve 80 diğer kişi

Like Yorum Yap Paylaş



User Language
Prediction

English: 0.99
French : 0.95



Language
Identification

Turkish: 0.99



See Translation

FACEBOOK AI

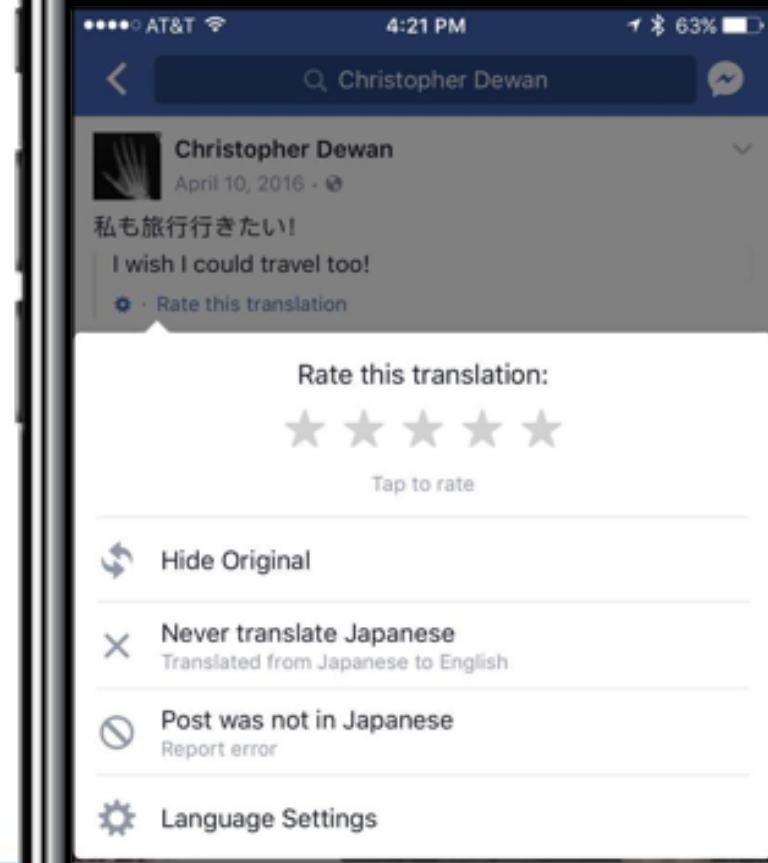
Georgia Tech



See Translation

FACEBOOK AI

Georgia Tech



Auto Translation

FACEBOOK AI

Georgia Tech

Summary

Translation Challenges: Low Resource



How to translate (Hausa)

ਮਾਰਾਂਗ ਮਾਰਿਅਰਿਕ)

ពੰਖੂਬੀ ਚੋਨੀ ਕੀਤੀ ਜਾਂਦੀ ਹੈ (Khmer)

NMT systems have a **steeper learning curve** with respect to the amount of training data , resulting in **worse quality in low-resource settings**, but better performance in high-resource settings.

Six Challenges for Neural Machine Translation - Koehn & Knowles (2017)

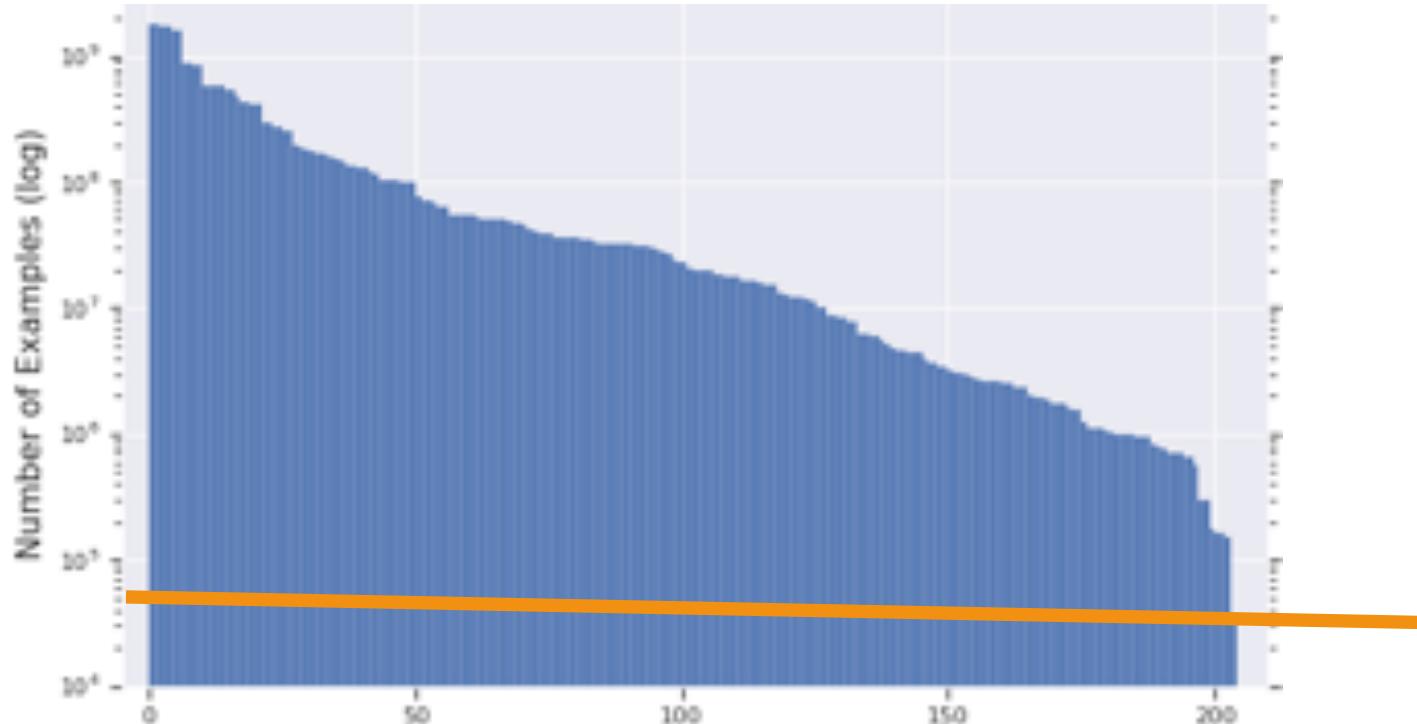


Chart credit: Siddhant et al., 2019

Evaluating the Cross-Lingual Effectiveness of Massively Multilingual Neural Machine Translation



Language
similarity



Domain

Evaluation



Language
similarity



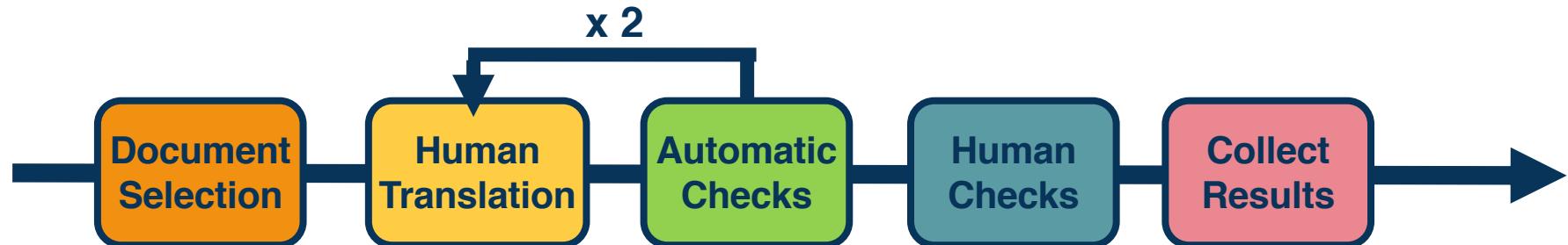
Domain



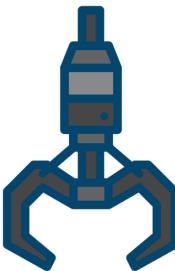
Evaluation

FLORES dataset:
Nepali-English, Sinhala-English





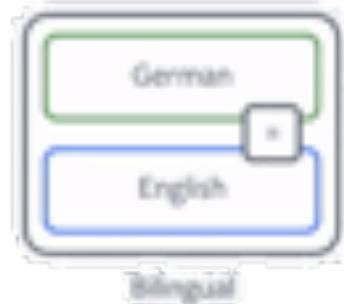
WIKIPEDIA
The Free Encyclopedia



Wikipedia logos by Wikimedia (CC-BY-SA)

Multilingual Training





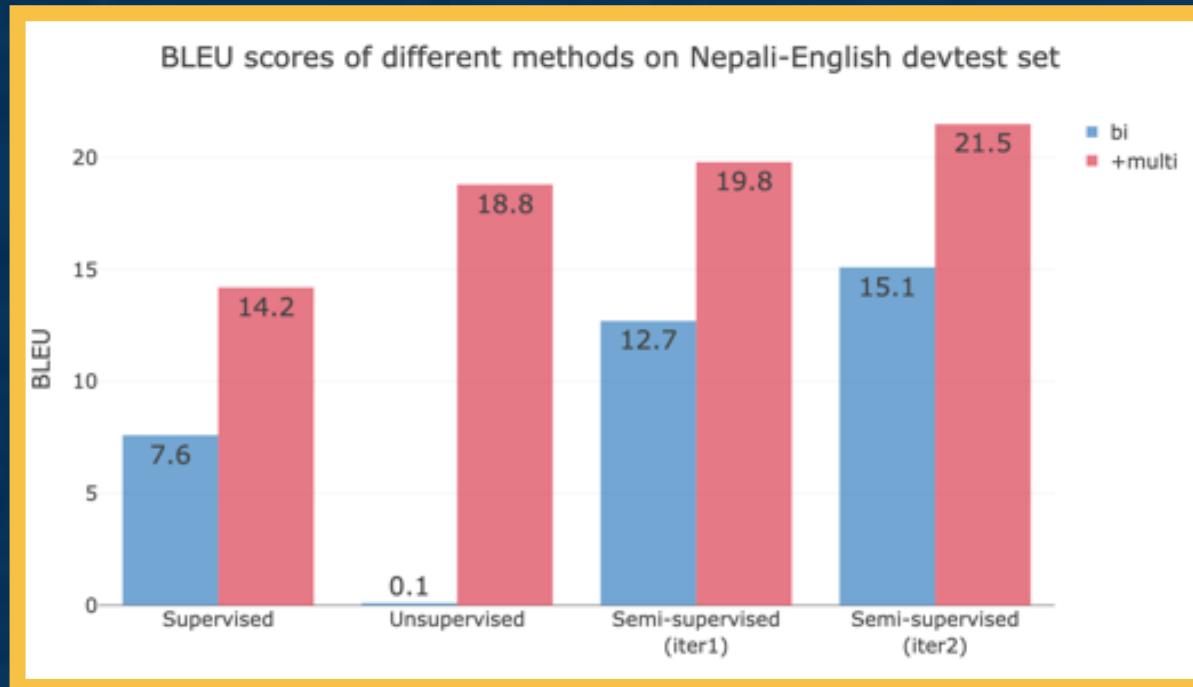


Iterative Backtranslation

FACEBOOK AI

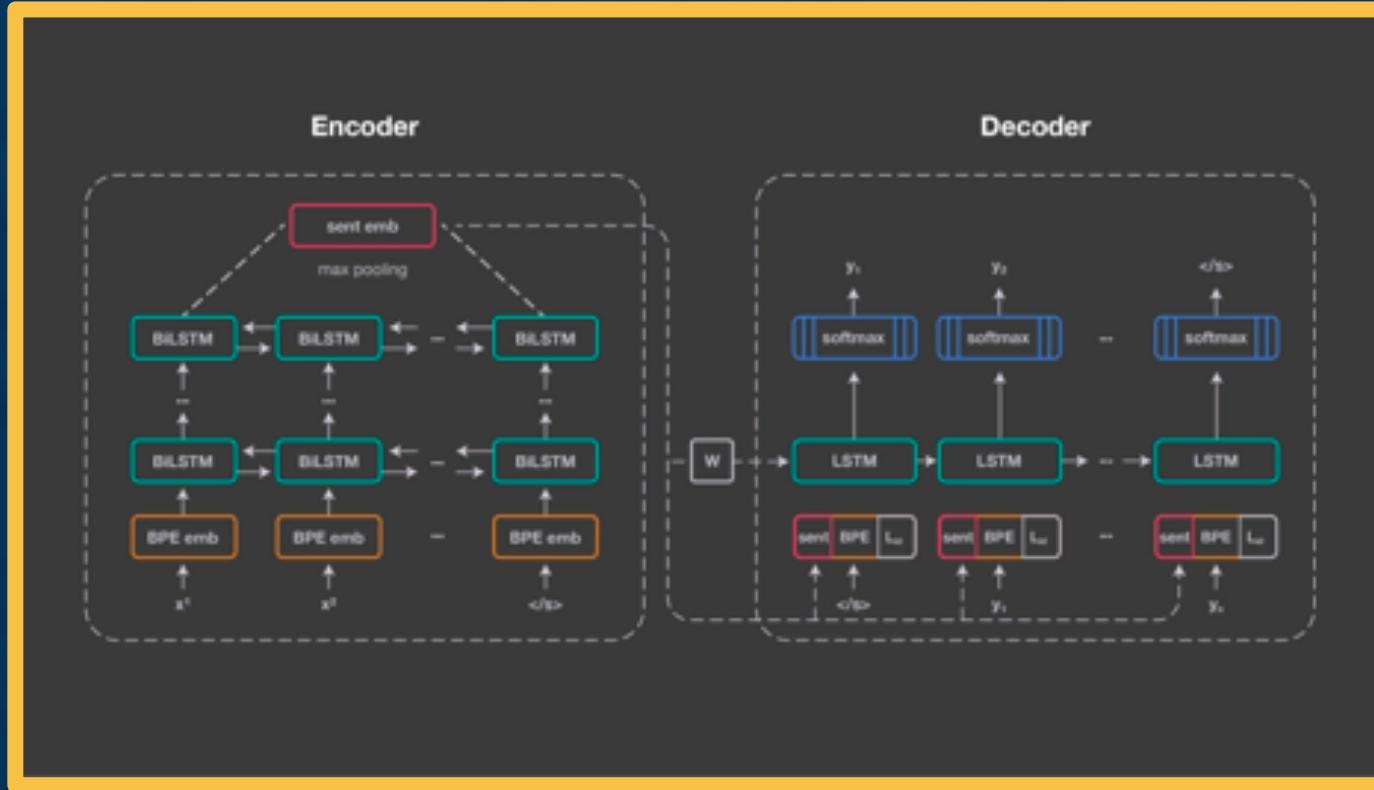
Georgia Tech²¹

Results

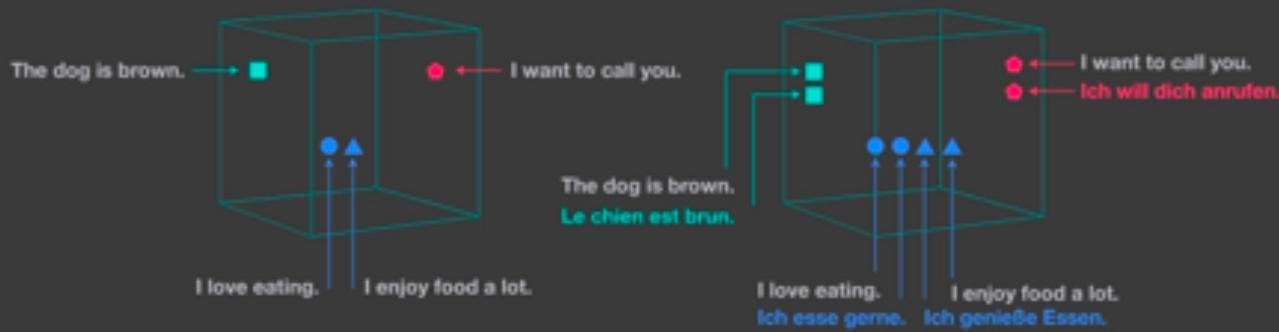


BLEU: precision-based metric that measures n-gram overlap with a human reference

Language Agnostic Sentence Representations



Language Agnostic Sentence Representations



Web-scale Mining

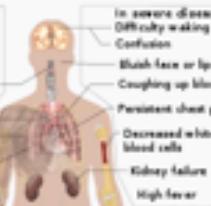
Corpora

WikiMatrix

CCMatrix

Causas	SARS-CoV-2
Factores de riesgo	hipertensión arterial obesidad diabetes mellitus enfermedad coronaria inmunodeficiencia Hipercolesterolemia enfermedad pulmonar obstructiva crónica enfermedad autoinmune Hiperglucemia asma aguda severa fumar tabaco asma
Diagnóstico	RT-PCR cuantitativa, imágenes de resonancia magnética axial computarizada
Prevención	Ver lista ²
Tratamiento	tratamiento sintomático antiviral monoterapia

Spanish

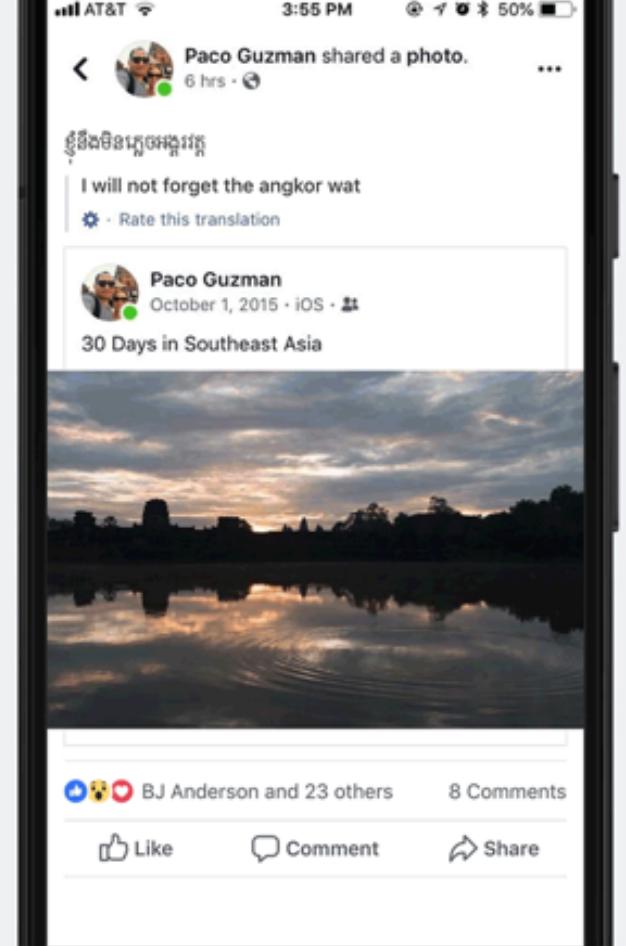
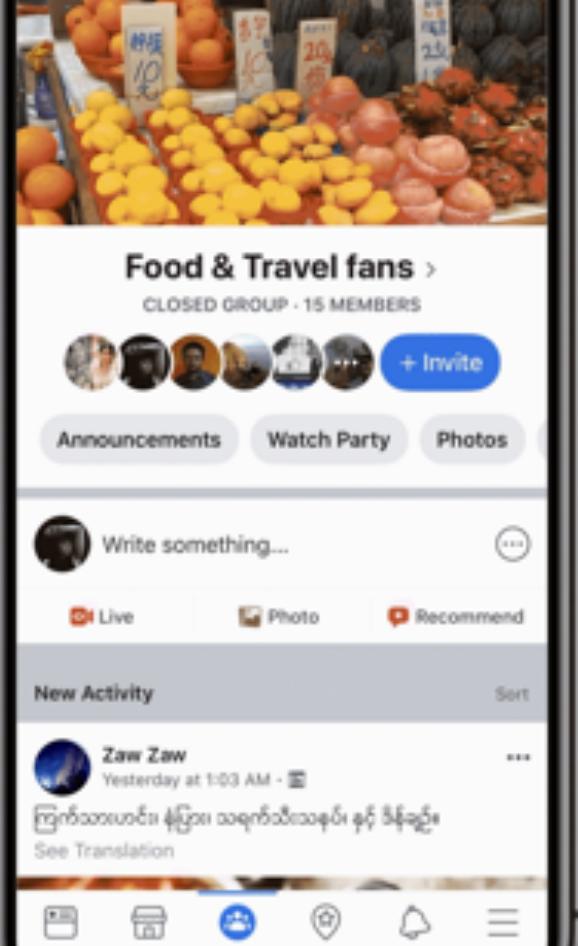
Other names	• Coronavirus • Corona • COVID • 2019-nCoV acute respiratory disease • Novel coronavirus pneumonia ^{[1][2]} • Severe pneumonia with novel pathogens ^[3]
Common symptoms	Fever: 83-93% Loss of appetite: 65-84% Fatigue: 44-70% Loss of smell: 15 to 30% Shortness of breath: 21-40% Cough: 53-82% Coughing up sputum: 26-33% Muscle aches and pain: 13-35%
	In severe disease: Difficulty walking Confusion Bluish face or lips Coughing up blood Persistent chest pain Decreased white blood cells Kidney failure High fever
Symptoms of COVID-19	
Pronunciation	/əˈkroʊnɪ.vərəs ˈdɪz.ɪz/ /koʊvɪdnaʊn.lɪn, ˈkoʊvɪd-/ ^[4]
Specialty	Infectious disease
Symptoms	Fever, cough, fatigue, shortness of breath, loss of smell; sometimes no symptoms at all ^{[5][6]}
Complications	Pneumonia, viral sepsis, acute respiratory distress syndrome, kidney failure, cytokine release syndrome
Usual onset	2-14 days (typically 5) from infection
Causes	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
Risk factors	Travel, viral exposure
Diagnostic method	RT-PCR testing, CT scan

English

FACEBOOK AI

Georgia Tech

ORIGINAL	BEFORE (USEFUL)	AFTER (GOOD)	HUMAN
સહરાની અન્ધું અન્ધું લાયકી કા અન્ધું અન્ધું અન્ધું રેન્ટોપ્પુર્ણું દિનું દિનું અન્ધું બાળ બાળ પ્રદૂષણકા.	Sahara is a very large area of land that is tongue to Egypt from the Atlantic border to Egypt.	Sahara is a very large land which runs from the Atlantic border to Egypt.	The Sahara is a very large expanse of land stretching from the Atlantic Rim to Egypt.
યગરની અન્ધું દર ગનીય અન્ધું કોઈ અન્ધું બાળ દુઃખ અન્ધું વિશેલ્પ વિશેલ્પ અન્ધું કરતું કા દેખકા.	Yogurt is one of the most commonly added to the spiciness of the spiciness.	Yogurt is a thing that will be added to food to reduce the strength.	Yogurt is a common addition to meals, as a way of tempering spiciness.
થોરે કોરેલા પારવિએરહરૂમા, એક સામાન્ય ભોજન ચામલ ર સંબંધિરૂ હુન્દન્દા।	In a lot of keila family, there are a normal meal chamal and vegetables.	In many kerala families, a common meal is rice and vegetables.	In most Kerala households, a typical meal consists of rice , and vegetables.
તોરીકો તેલ પારમ્પરાકિ રૂપલે ખાન તેથાર ગર્નમા પ્રયુક્ત હુન્દા।	Torico Oil is used to prepare khan traditional.	The Oil is traditionally used to prepare food.	Mustard oil is traditionally used in cooking food.



Bring Support to Low-Resource Languages

FACEBOOK AI

Georgia Tech

Summary

Rethinking Translation Quality

When Things Go Wrong

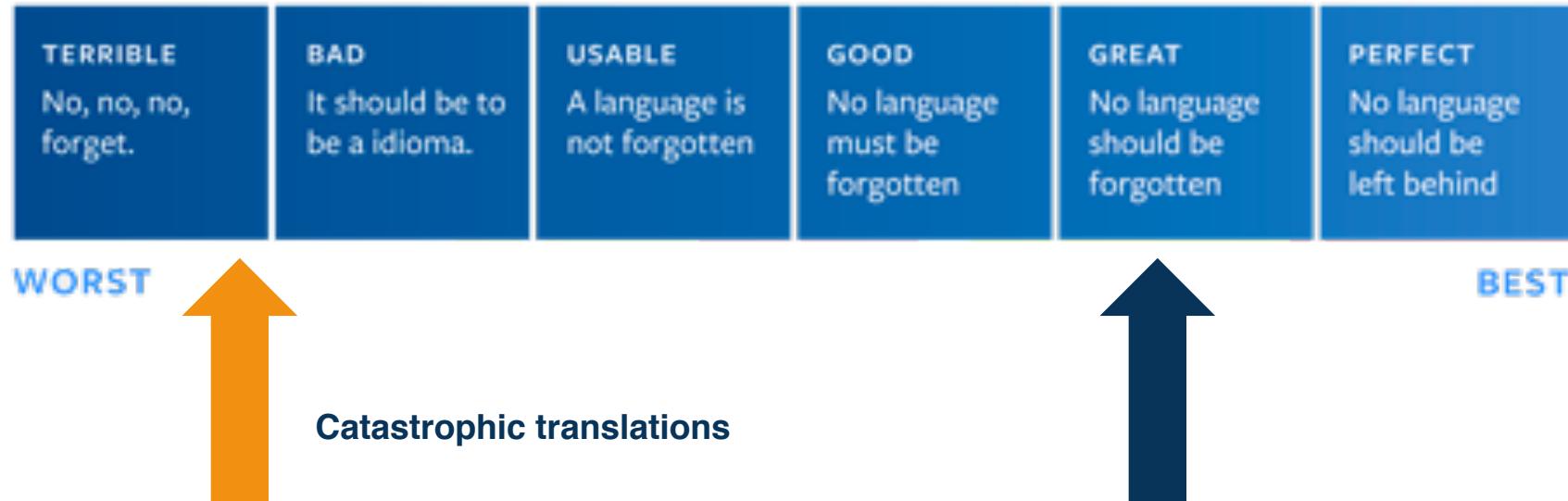
When we translated *good morning* into “attack them”, leading into an arrest.

When names of public figures have been misunderstood, resulting in offensive translations.



SOURCE

Ningún idioma debe ser olvidado



Quality is Not Linear

FACEBOOK AI

Georgia Tech

What Throws Users Off? Catastrophic Translations

Bad Named Entity translation

(El Centro, CA -> The Center,
CA)

Wrong pronouns (*it, he* instead
of *she*)

False profanities (*Mr sh*thole*)

Implausible translations (a water
treadmill)

Introduction of violent terms
("attack them")



Future: Translations Centered Around User-Experience

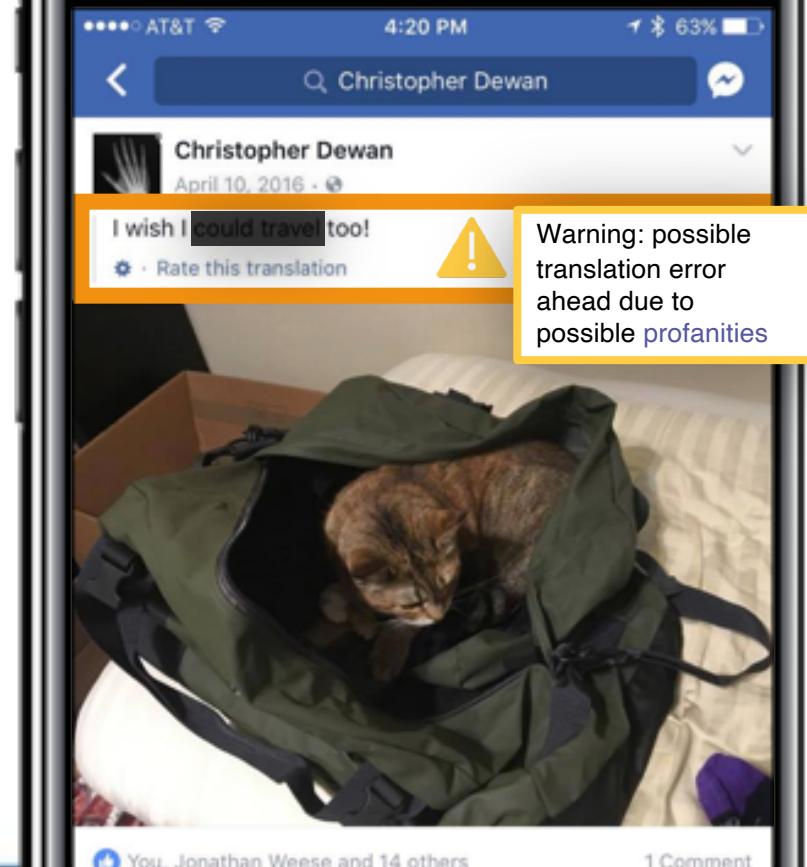
Catastrophic-aware translation systems

Self Explanatory

Quality metrics that account for critical mistakes

AI Systems with more transparency and control





Hypothetical UI Experience

FACEBOOK AI

Georgia Tech

Summary