- Question Answering systems (Q.A.) +lab (Q.A. finetune BERT model)
- Multilingual and Multimodal Q.A. +lab (Q.A. test BERT models)
- Information Retrieval (I.R.) +lab (I.R. train and test BM25 model)
- Open Domain Q.A. +assingment (Open Domain Q.A)

- Question Answering systems (Q.A.) +lab
- Multilingual and Multimodal Q.A. +lab
- Information Retrieval (I.R.) +lab
- Open Domain Q.A. +assignment

- Multilingual and Multimodal Q.A.
  - Introduction
  - Multilingual Dataset & Strategies
  - Multimodal Q.A. (small overview)

## What is multilingual question answering?



The goal of multilingual question answering is to build systems that automatically answer questions posed by humans in 'any' language

## What is multilingual question answering?



The goal of multilingual question answering is to build systems that automatically answer questions posed by humans in 'any' language

However, most of the datasets in NLP are in English... SQuAD, QuAC....

Solution?

## What is multilingual question answering?



The goal of multilingual question answering is to build systems that automatically answer questions posed by humans in 'any' language

However, most of the datasets in NLP are in English... SQuAD, QuAC....

- Translate Train (MT + alingment)
- Translate Test (MT + alingment)
- Zero-Shot (CrossLingual pretrained L.M.)

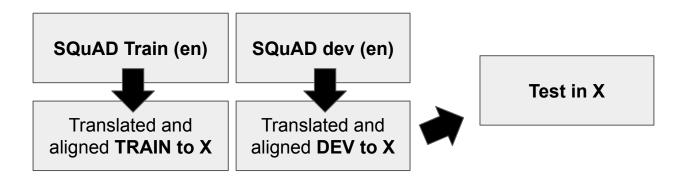
## Translate Train (and Dev)



**Model:** monolingual BERT model trained in X

**Train/Dev:** Translate Question, Passage and Answer using M.T. + align the answer

Test: -



#### **Translate Test**



**Model:** monolingual BERT model trained in English

Train/Dev: -

Test: Translate Question, Passage and tag Answer + align the answer back to X



#### Zero Shot!



**Model:** multilingual BERT model (104 languages)

Train/Dev: -

Test: -



We don't need M.T. and alignment!

- Multilingual and Multimodal Q.A.
  - Introduction
    - Multilingual pre-trained L.M.
  - Multilingual Dataset & Strategies
  - Multimodal Q.A. (small overview)

## MLQA Multilingual Question Answering Multilingual pre-trained L.M.s



#### mBERT 104 languages

- Wikipedia (Low-resource languages are upweighted by some factor)
- MLM+NSP

**XLM** (BERT arch.) 15 languages

- Wikipedia + Parallel Corpora
- MLM+TLM

#### XLM-Roberta 100 languages

- Common Crawl
- MLM

## MLQA Multilingual Question Answering Multilingual pre-trained L.M.s



#### mBERT 104 languages

- Wikipedia (Low-resource languages are upweighted by some factor)
- MLM+NSP

#### **XLM** (BERT arch.) 15 languages

- Wikipedia + Parallel Corpora
- MLM+TLM

#### XLM-Roberta 100 languages

- Common Crawl
- MLM
- XLM-T RoBERTuito twitterXLMR (twitter)
- Check HuggingFace Models :)
- ..

## MLQA Multilingual Question Answering Multilingual pre-trained L.M.s

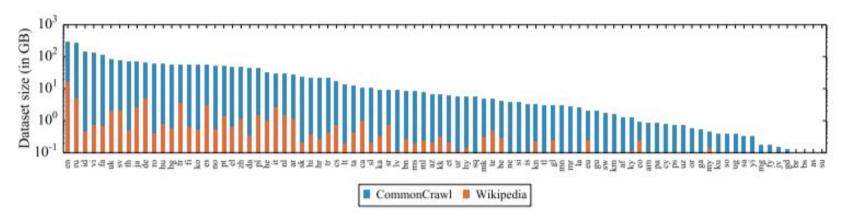
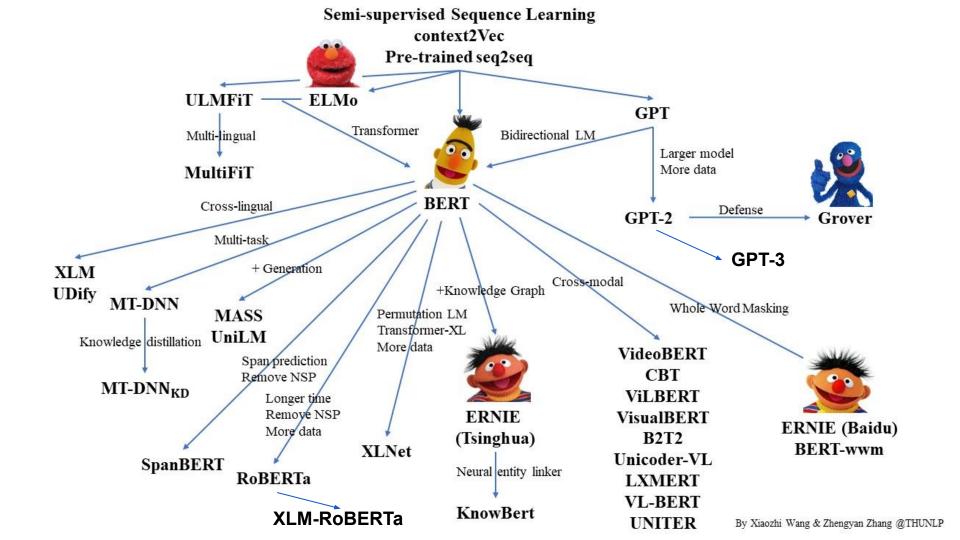
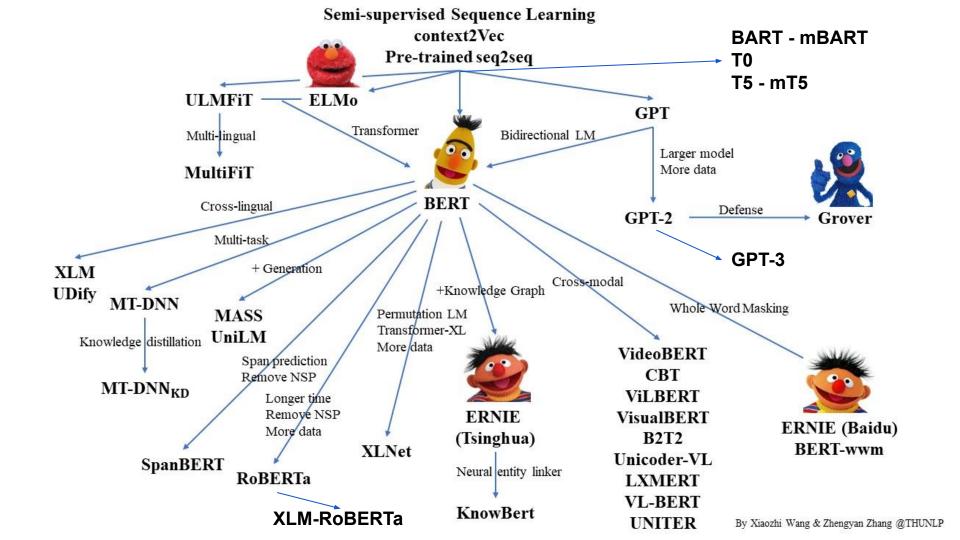


Figure 1: Amount of data in GiB (log-scale) for the 88 languages that appear in both the Wiki-100 corpus used for mBERT and XLM-100, and the CC-100 used for XLM-R. CC-100 increases the amount of data by several orders of magnitude, in particular for low-resource languages.





- Multilingual and Multimodal Q.A.
  - Introduction
  - Multilingual Dataset & Strategies
    - MLQA Multilingual Question Answering dataset
  - Multimodal Q.A. (small overview)

- MLQA (MultiLingual Question Answering) is a benchmark dataset for evaluating cross-lingual question answering performance.
- MLQA consists of over 5K extractive QA instances (12K in English) in SQuAD format in seven languages - English, Arabic, German, Spanish, Hindi, Vietnamese and Simplified Chinese.
- MLQA is highly parallel, with QA instances parallel between 4 different languages on average.
- Only dev and test!

https://github.com/facebookresearch/MLQA

Patrick Lewis, Barlas Oğuz, Ruty Rinott, Sebastian Riedel, Holger Schwenk, MLQA: Evaluating Cross-lingual Extractive Question Answerin (2020)

Fn

During what time period did the Angles migrate to Great Britain?

The name "England" is derived from the Old English name Englaland [...] The Angles were one of the Germanic tribes that settled in Great Britain during the Early Middle Ages. [...] The Welsh name for the English language is "Saesneg"

De

Während welcher Zeitperiode migrierten die Angeln nach Großbritannien?

Der Name England leitet sich vom altenglischen Wort Engaland [...] Die Angeln waren ein germanischer Stamm, der das Land im *Frühmittelalter* besiedelte. [...] ein Verweis auf die weißen Klippen von Dover.

Ar

في أي حقبة زمنية هاجر الأنجل إلى بريطانيا العظمى؟

والتي تعني "أرض الأنجل". والأنجل كانت واحدة Englaland يشتق اسم "الجلنرا" من الكلمة الإنجليزية القديمة من القبائل الجرمائية التي استقرت في الجلنرا خلالم *أوائل العصور الوسطى.* [...] وقد سماها العرب قديما الانكناء

٧

Trong khoảng thời gian nào người Angles di cư đến Anh?

Tên gọi của Anh trong tiếng Việt bắt ngườn từ tiếng Trung. [...] Người Angle là một trong những bộ tộc German định cư tại Anh trong *Thời đầu Trung Cổ*. [...] dường như nó liên quan tới phong tục gọi người German tại Anh là Angli Saxones hay Anh - Sachsen.

En

What are the names given to the campuses on the east side of the land the university sits on?

The campus is in the residential area of Westwood [...] The campus is informally divided into *North Campus and South Campus*, which are both on the eastern half of the university's land. [...] The campus includes [...] a mix of architectural styles.

Es

¿Cuáles son los nombres dados a los campus ubicados en el lado este del recinto donde se encuentra la universidad?

El campus incluye [...] una mezcla de estilos arquitectónicos. Informalmente está dividido en *Campus Norte y Campus Sur*, ambos localizados en la parte este del terreno que posee la universidad. [...] El Campus Sur está enfocado en la ciencias físicas [...] y el Centro Médico Ronald Reagan de UCLA.

Zh

位于大学占地东半部的校园名称是什么?

整个校园被不正式地分为**南北两个校园**,这两个校园都位于大学占地的东半部。 北校园是原校园的中心,建筑以义大利文艺复兴时代建筑闻名,其中的包威尔图 书馆(Powell Library)成为好莱坞电影的最佳拍摄场景。[...] 这个广场曾在许多 电影中出现。

ш

विश्वविद्यालय जहाँ स्थित है, उसके पूर्वी दिशा में बने परिसरों को क्या नाम दिया गया है?

जब 1919 में यूसीएलए ने अपना नया परिसर खोला, तब इसमें चार इमारतें थी। [...] परिसर अनौपचारिक रूप से उत्तरी परिसर और दक्षिणी परिसर में विभाजित है, जो दोनों विश्वविद्यालय की जमीन के पूर्वी हिस्से में स्थित हैं। [...] दक्षिणी परिसर में भौतिक विज्ञानं, जीव विज्ञान, इंजीनियरिंग, मनोविज्ञान, गणितीय विज्ञान, सभी स्वास्थ्य से संबंधित क्षेत्र और युएलसीए मेडिकल सेंटर स्थित है।

(a)

(b)

The MLQA paper presents several **baselines** (fscores) for zero-shot experiments on MLQA, with training QA data taken from SQuAD V1.1, and using the MLQA English development set for early stopping.

## Zero-Shot!

Model F1 Score	en	es	de	ar	hi	vi	zh
BERT-Large	80.2	12	-	20	2	12	
Multilingual-BERT	77.7	64.3	57.9	45.7	43.8	57.1	57.5
XLM	74.9	68.0	62.2	54.8	48.8	61.4	61.1
Translate-test BERT-L	-	65.4	57.9	33.6	23.8	58.2	44.2
Translate-train M-BERT	-	53.9	62.0	51.8	55.0	62.0	61.4
Translate-train XLM		65.2	61.4	54.0	50.7	59.3	59.8

The MLQA paper presents several **baselines** (fscores) for zero-shot experiments on MLQA, with training QA data taken from SQuAD V1.1, and using the MLQA English development set for early stopping.

Zero-Shot!

Model F1 Score	en	es	de	ar	hi	vi	zh
BERT-Large	80.2	12	-	20	2	12	
Multilingual-BERT	77.7	64.3	57.9	45.7	43.8	57.1	57.5
XLM	74.9	68.0	62.2	54.8	48.8	61.4	61.1
Translate-test BERT-L	-	65.4	57.9	33.6	23.8	58.2	44.2
Translate-train M-BERT	-	53.9	62.0	51.8	55.0	62.0	61.4
Translate-train XLM	-	65.2	61.4	54.0	50.7	59.3	59.8

x100~ Multilingual Models: mBERT, XLM, XLMR. mBART....

x3 Multilingual Models: IXAmBERT https://www.ehu.eus/ehu sfera/ixa/2020/09/30/ixa mbert-good-news-for-lan guages-with-few-resourc es/

#### **LREC2020**

The MLQA paper presents several **baselines** (fscores) for zero-shot experiments on MLQA, with training QA data taken from SQuAD V1.1, and using the MLQA English development set for early stopping.

Zero-Shot!

Model F1 Score	en	es	de	ar	hi	vi	zh
BERT-Large	80.2	12	-	20	2	12	
Multilingual-BERT	77.7	64.3	57.9	45.7	43.8	57.1	57.5
XLM	74.9	68.0	62.2	54.8	48.8	61.4	61.1
Translate-test BERT-L	-	65.4	57.9	33.6	23.8	58.2	44.2
Translate-train M-BERT	-	53.9	62.0	51.8	55.0	62.0	61.4
Translate-train XLM	-	65.2	61.4	54.0	50.7	59.3	59.8

x100~ Multilingual Models: mBERT, XLM, XLMR, mBART....

x3 Multilingual Models: IXAmBERT https://www.ehu.eus/ehu sfera/ixa/2020/09/30/ixa mbert-good-news-for-lan guages-with-few-resourc es/

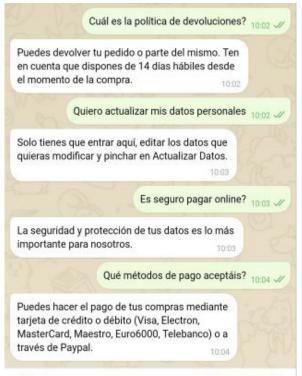
translate train > zero shot > translate test ??? not always

# MLQA Multilingual Question Answering Best practices for multilingual model training

MLQA is intended to be an evaluation corpus. **Please limit evaluations on the test set to an absolute minimum to prevent overfitting.** There is a development dataset split which can be used for running intermediate evaluations during model development.

If you are running experiments on MLQA, it is important that you clearly state your experimental settings. If you are performing zero-shot experiments, you should only use the development data in the language you are training on. Using MLQA dev data for particular target language to tune performance in that target language is a valid research direction, but is not strictly zero-shot and you should make sure that you explicitly state how you use the development data to ensure fair comparison in the future.

## Use case: Q.A. over F.A.Q. for spanish



QA system that replies to questions using FAQ repositories, in any language.

Question Answering











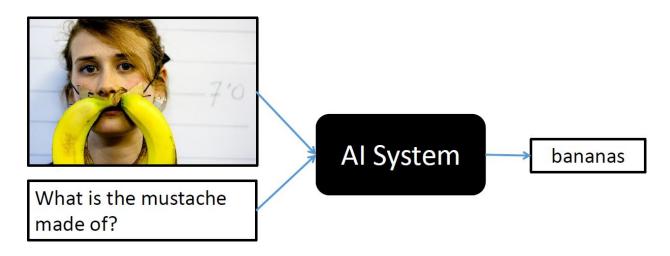


#### Use case: Q.A. over F.A.Q. for spanish

- Train mBERT in SQuAD, QuAC, DoQA...
  - Don't use previous history and ignore no-answers for QuAC ~~~ SQuAD
  - Train SQuAD + QuAC
- Test in Out of Domain <u>BarkBox</u> data (dog suplies)
  - Best: only QuAC training data (long answers were preferred)
  - Preferred by most users
- Finetune **mBERT** language model with F.A.Q question (in domain) and test in <u>BarkBox</u> data
  - ~ Same results

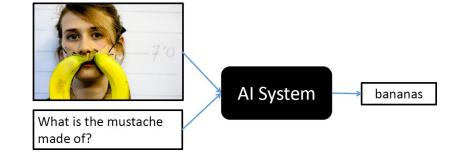
- Multilingual and Multimodal Q.A.
  - Introduction
  - Multilingual Dataset & Strategies
  - Multimodal Q.A. (small overview)

#### Visual Q.A. challenge



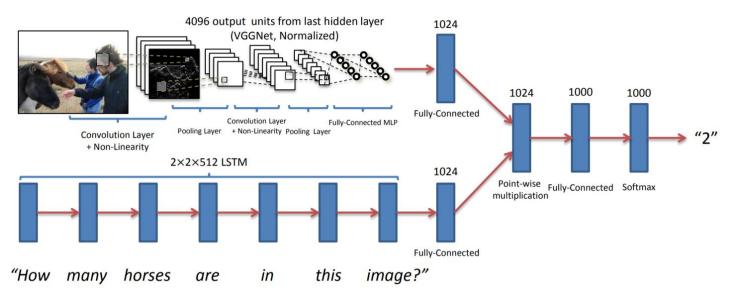
- Questions against images, not text. <a href="https://visualga.org/challenge.html">https://visualga.org/challenge.html</a>
- Given an image and a natural language question about the image, the task is to provide an accurate natural language answer (pick from a predefined list of possible answers).
  - 250K images and 1.1M questions
  - LeaderBoard <u>2020</u> (+papers)

### Visual Q.A. challenge



<u>Aishwarya Agrawal</u>, <u>Jiasen Lu</u>, <u>Stanislaw Antol</u>, <u>Margaret Mitchell</u>, <u>C</u>. <u>Lawrence Zitnick</u>, <u>Dhruv Batra</u>, <u>Devi Parikh</u>

<u>VQA: Visual Question Answering 2016</u>



- Multilingual and Multimodal Q.A.
  - Introduction
  - Multilingual Dataset & Strategies
  - Multimodal Q.A. (small overview)
    - Demos

## Visual Q.A. demos

- <u>VQA</u>
- Visual Chat Bot
- Visual Chat Bot demo video

#### Thanks!

ander.barrena@ehu.eus @4nderB

#### Lab Time! <a href="http://ixa2.si.ehu.eus/~jibloleo/nlpapp2">http://ixa2.si.ehu.eus/~jibloleo/nlpapp2</a>

#### • [7. Question Answering with a fine-tuned BERT]

- Edited from Chris McCormick notebook
- Code:
  - Encode question, answer pairs
  - Segment ids
  - Get BERT output: bert-large-uncased-whole-word-masking-finetuned-squad
  - Get argmax for start and end tokens
  - Extract answer from original text
    - Detokenize BPE
  - Spam Scores
- Assignments:
  - Improve over argmax for start end tokens
  - Print answer score
  - Zero-shot model (check Huggingface models) search for multilingual & Q.A.
    - https://huggingface.co/models