

Language Modelling with Pixels

Explorations of pixel-based encoding of language

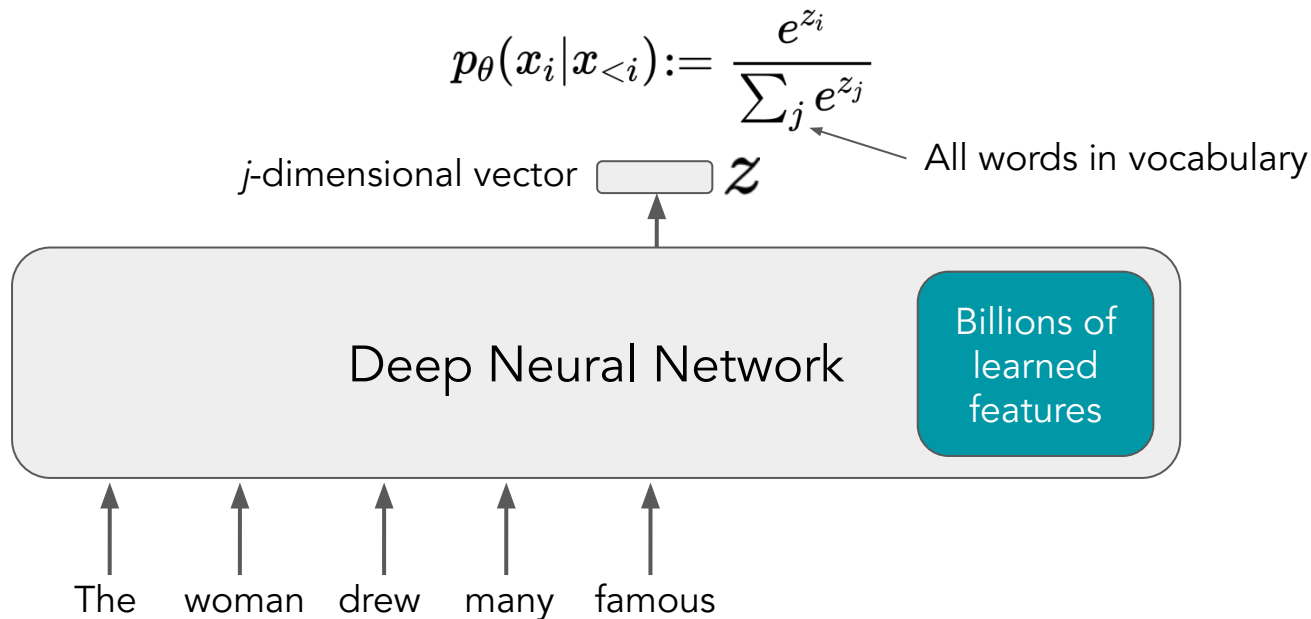


Jonas F. Lotz

Department of Computer Science, University of Copenhagen

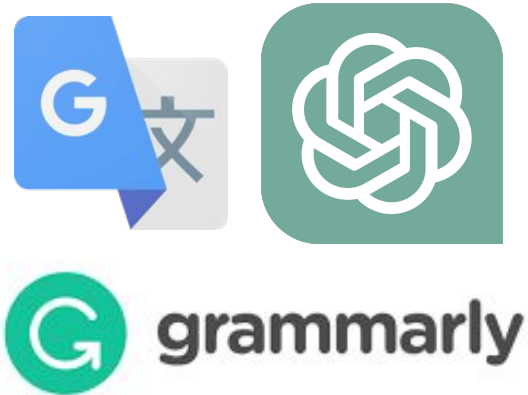


Large Language Models?

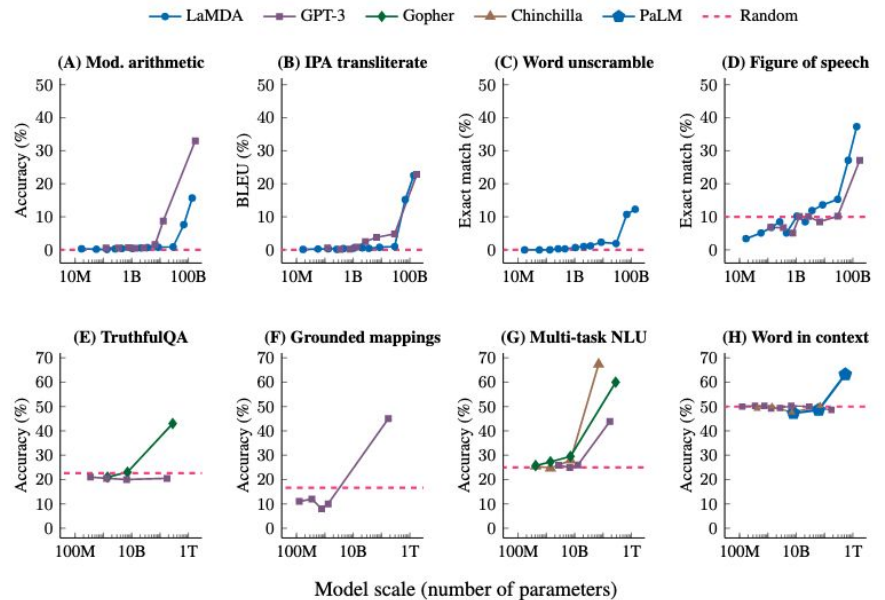
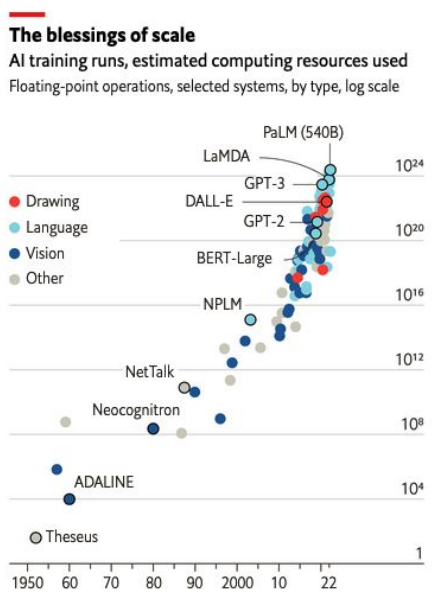


Large Language Models

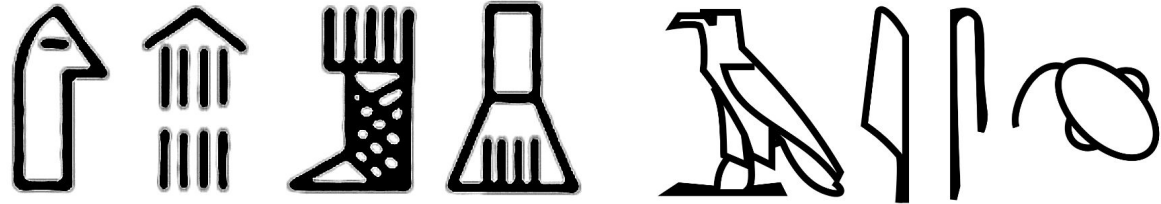
- Spelling and grammar checking
- Machine translation
- Web search
- Text prediction
- Topic modelling
- Chatbots



In the Era of Scale



What's left?

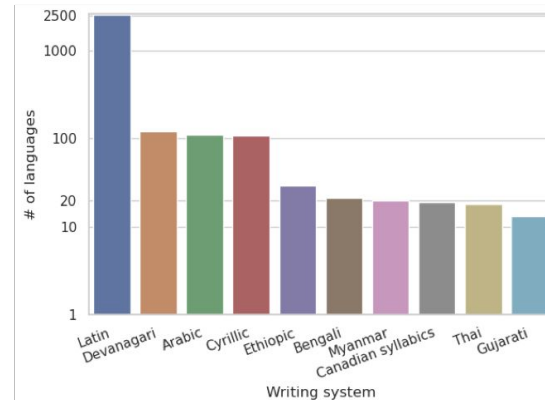
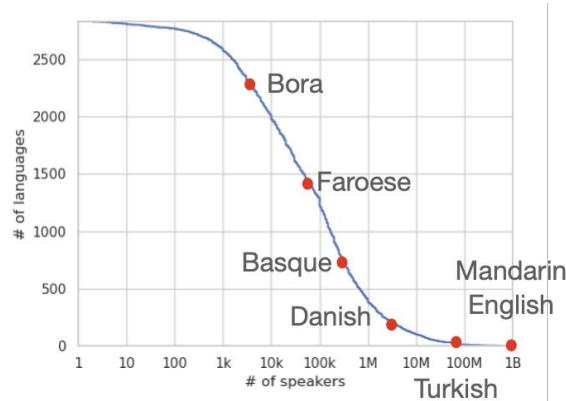


天地玄黃 𠄎 𠄎 𠄎 𠄎

ABCD अ इ उ ण्।

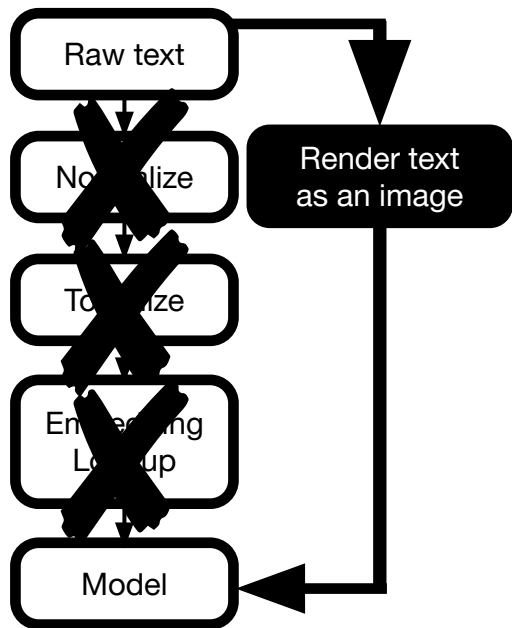
NLP for All Written Languages

- There are 7,000 spoken languages, of which 3,000 are written
 - There is at least 400 languages with >1M speakers
- But NLP only covers 100 languages ([van Esch+ LREC22](#))
 - Lack of technological inclusion for billions of people



Today: Pixel-based Language Modelling

- Key insight: treat language processing as visual processing



Søren Kierkegaard (d. 1855) was a Golden Age philosopher

S	ø	r	e	n	K	i	e	r	k	e	g	a	a	r	d	(d	.	1	8	5	5)	w	a	s	a	G	o	l	d	e	n	A	g	e	p	h	i	l	s	o	p	h	e	r
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

ኢትዮጵያ አፍሪካ ውስጥ ናት

ኢ	ት	ዮ	ጵ	ያ	አ	ፍ	ሪ	ካ	ው	ስ	ጥ	ና	ት
---	---	---	---	---	---	---	---	---	---	---	---	---	---

A new type of generative model

Penguins are designed to be streamlined and hydrodynamic, so **having thin legs** would add expanding. Having short legs with webbed feet to act like rudders, helps to give them that the le do-like figure didn't compare bird anatomy with humans, we would see something **is** peculiar. By taking a look at the side-by-side image in Figure 1, you can see how their leg **bones are** to ours. What most people mistake for **knees** are actually the **anatomies of birds**. This **gives a conclusion** that bird knees bend opposite of ours. The knees are actually tucked up inside the **boxes** of the **bird**. So how does this look inside the penguin? In the **images** below, you can see boxes surrounding the penguins' knees.

Penguins are designed to be streamlined and hydrodynamic, so **having long legs** would add expanding. Having short legs with webbed feet to act like rudders, helps to give them that **these** do-like figures **is** to compare bird anatomy with humans, we would see something **is** peculiar. By taking a look at the side-by-side image in Figure 1, you can see how their leg **bones are** to ours. What most people mistake for **knees** are actually the **anatomies of birds**. This **gives the clusion** that bird knees bend opposite of ours. The knees are actually tucked up inside the **boxes** of the **bird**. So how does this look inside of a penguin? In the **images** below, you can see boxes surrounding the penguins' knees.

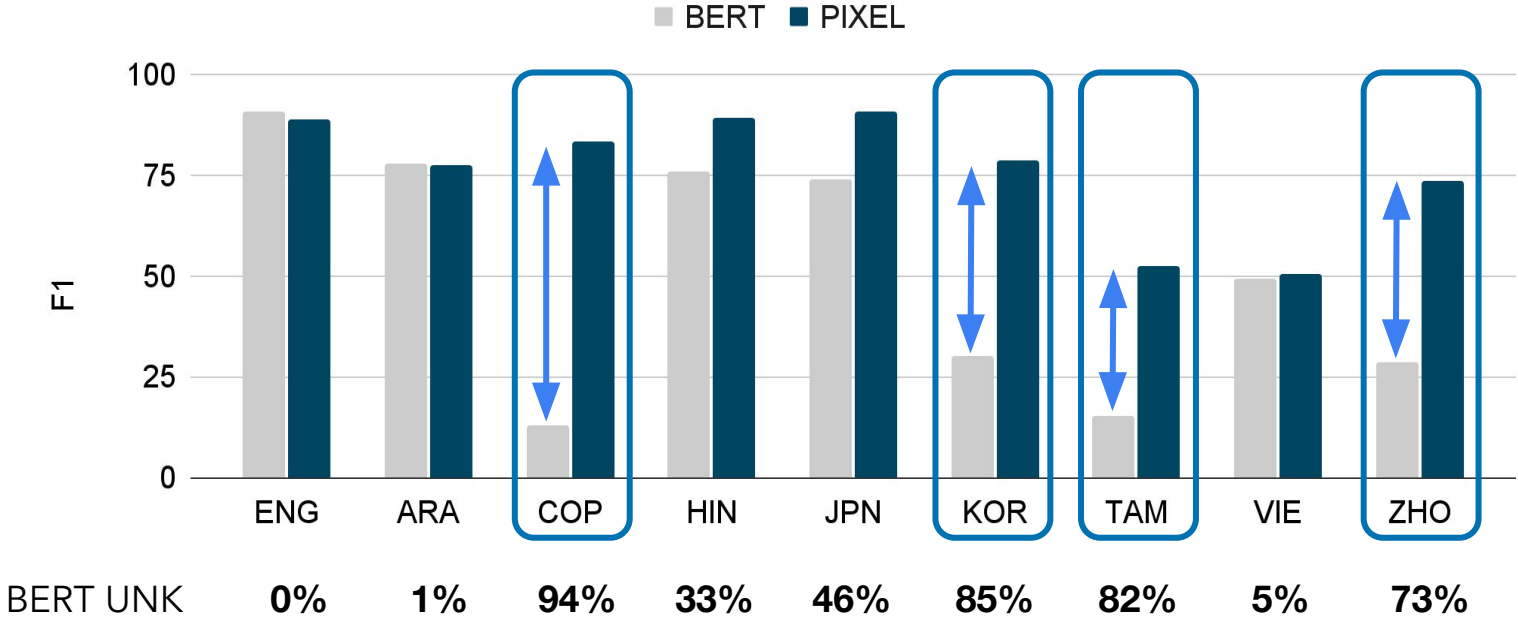
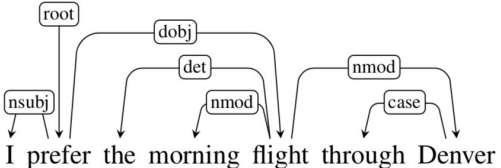
Penguins are designed to be streamlined and hydrodynamic, so **having long legs** would add expanding. Having short legs with webbed feet to act like rudders, helps to give them that **these** do-like figure. **If** we compare bird anatomy with humans, we would see something **is** peculiar. By taking a look at the side-by-side image in Figure 1, you can see how their leg **bones are** to ours. What most people mistake for **knees** are actually the **anatomies of birds**. This **gives the illusion** that bird knees bend opposite of ours. The knees are actually tucked up inside the **boxes** of the **bird**. So how does this look inside of a penguin? In the **images** below, you can see boxes surrounding the penguins' knees.

100K steps

500K steps

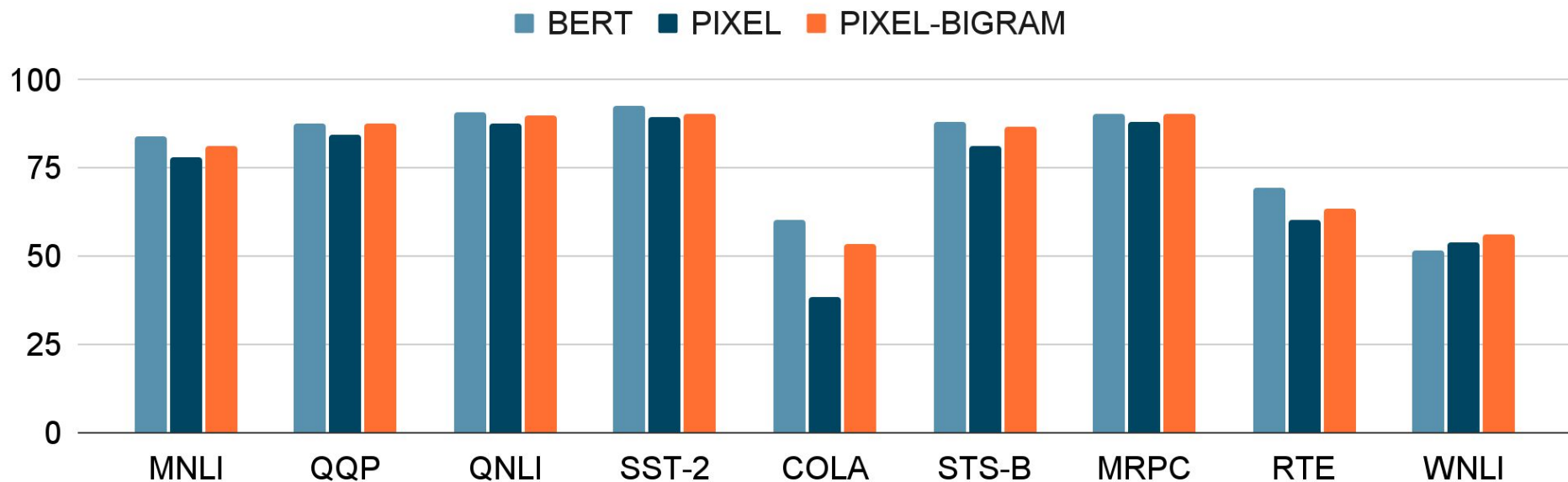
1M steps

Results: Dependency Parsing



PIXEL vastly outperforms BERT on unseen scripts

Exploring Text Rendering Strategies



Bigram text rendering produces much better models

Pretraining

- **English Dataset:** English Wikipedia and Books Corpus
- **Masking:** 25% Span Masking
- **Maximum sequence length:** 529 patches (16x8464 pixels)
- **Compute:** 8 x 40GB A100 GPUs for 8 days
- **Parameters:** 86M encoder + 26M decoder

There is only 0.05% non-English text in our pretraining data (estimated by Blevins and Zettlemoyer 2022)

The **Great Wall of China** (traditional Chinese: 萬里長城; simplified Chinese: 万里长城; pinyin: Wànlǐ Chángchéng)

Ongoing Work

- Improve sentence level reasoning tasks
 - Contrastive objective or multi-scale modelling
- Scale to multilingual pretraining
 - Better representations from reconstructing multiple scripts?
- Understanding cross-script transfer
 - What causes the PIXEL model transfer well to other scripts?