

# Revisiting N-Gram Models: Their Impact in Modern Neural Networks for HTR

Solène Tarride and Christopher Kermorvant

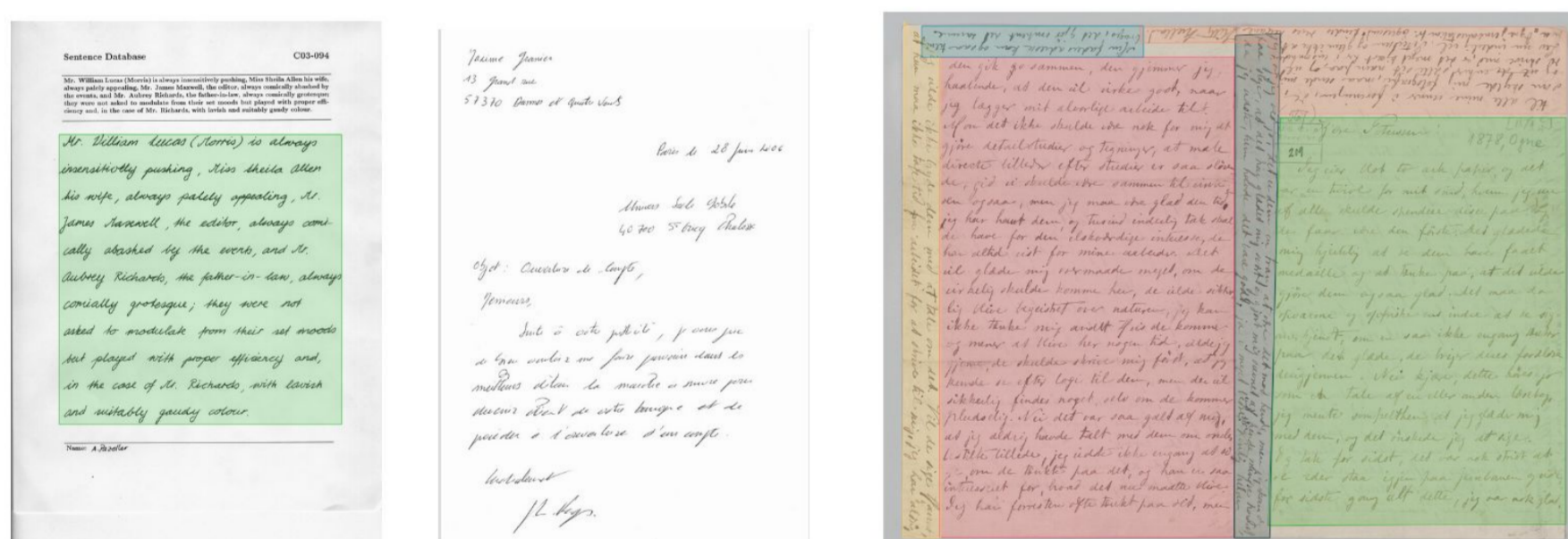
## Are n-gram models still useful with modern architectures?

- N-gram language models were once crucial to enhance performance in Automatic Text Recognition (ATR)
- Over time, explicit language models have been overshadowed by models with implicit language modeling capabilities (transformers)
- Integrating explicit n-gram models with transformers
  - Impact on transformer performance vs. CRNN-CTC?
  - Effect on inference speed?
- N-gram models are applicable at different text granularity
  - Which granularity is the most effective? Characters? Subwords? Words?

## Methodology and experiments

### Datasets

- IAM Lines & Pages (no header)
- RIMES Lines & Pages
- NorHand Lines & Paragraphs



### Models

- PyLaia
  - CRNN with CTC decoder
  - Trained on text-lines
- Document Attention Network (DAN)
  - CNN encoder with Transformer decoder
  - Trained on text-lines and on pages/paragraphs

### Explicit language modeling

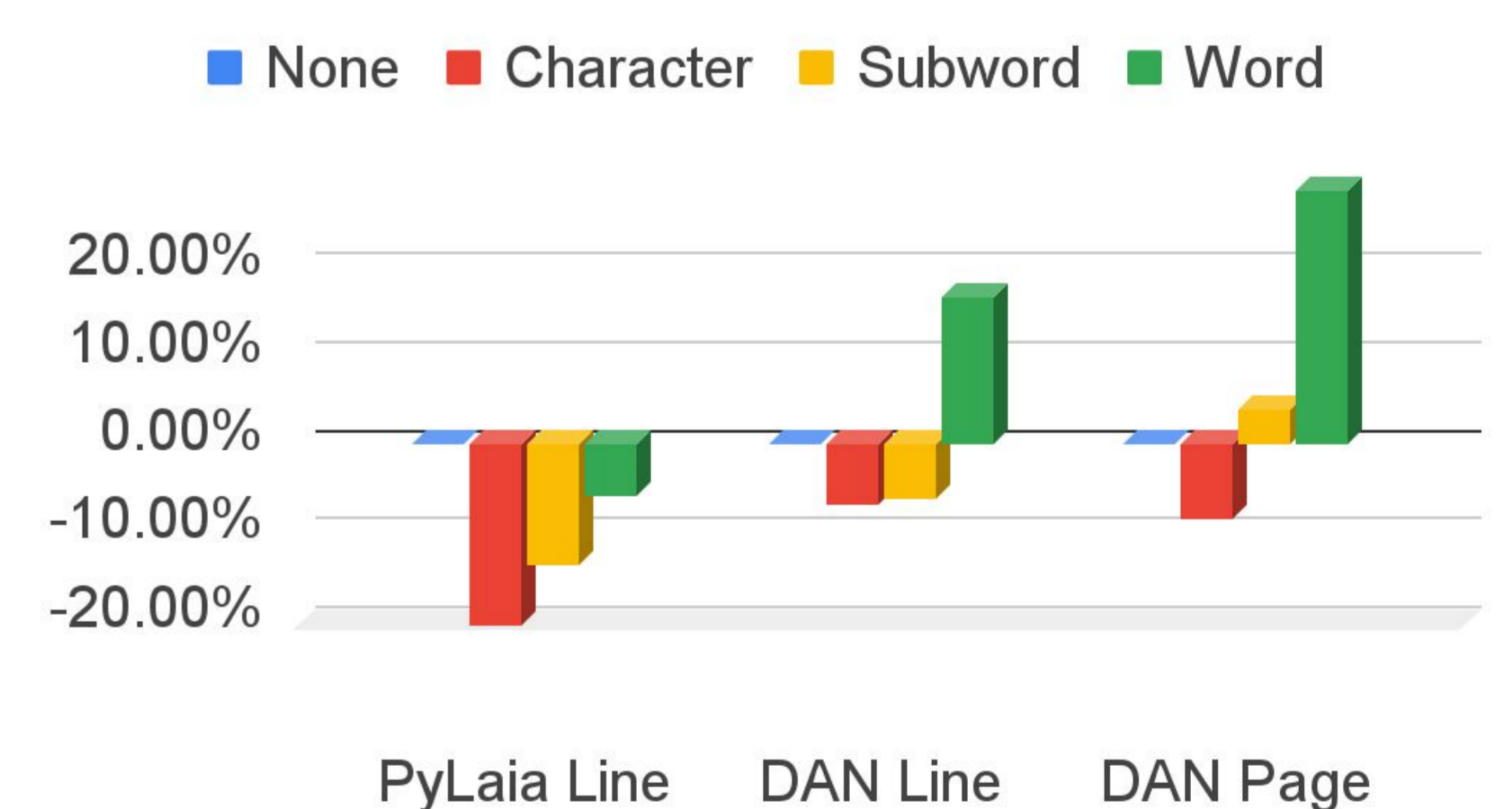
- N-gram language models are built with the KenLM library
- Tested with different tokenization strategies
  - Character (6-gram)
  - Subword (6-gram)
  - Word (3-word)

Level	Tokenized text
Character	T h e _ n u m e r i c a l l y _ l a r g e s t _ g r o u p
Subword	The _ numer ically _ large st _ gro up
Word	The _ numerically _ largest _ group

- We combined PyLaia and DAN with n-gram models thanks to Torchaudio's ctc\_decoder function

### Results

- Relative change in Word Error Rate (WER) averaged on all datasets
  - Character LM: -20% WER
  - Subword LM: -5% WER
  - Word LM: +13% WER



- Character n-gram models benefit PyLaia more than DAN:
  - 20% improvement for PyLaia
  - 8% improvement for DAN
- Language models impact speed: 10x slower for PyLaia, 1.4x for DAN

## TL;DR

- Are n-gram models still useful with modern architectures? **Yes!**
- N-gram models still improve performance for HTR :
  - 20% reduction in WER for PyLaia
  - 8% reduction in WER for DAN
- Granularity matters:
  - Character n-gram models are great
  - Subword n-gram models are effective
  - Word n-gram models are counterproductive
- Performance improves, but **at the cost of a higher inference time**

- Full paper:



- Contact information: Solène Tarride ([starride@tekliia.com](mailto:starride@tekliia.com))  
<https://tekliia.com/>  
[@\\_Teklia\\_](https://twitter.com/_Teklia_)