Ancestor-to-Creole Transfer is not a Walk in the Park
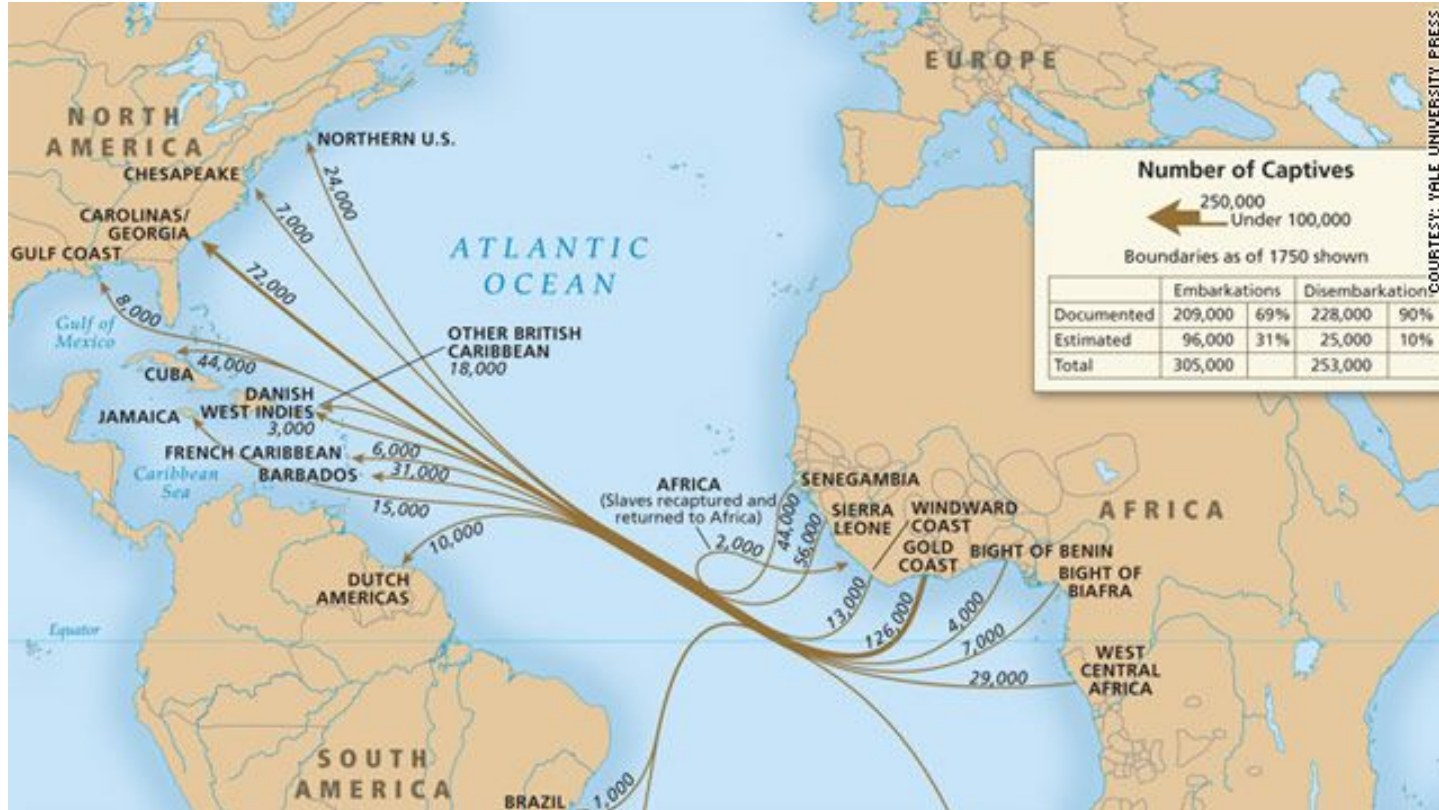
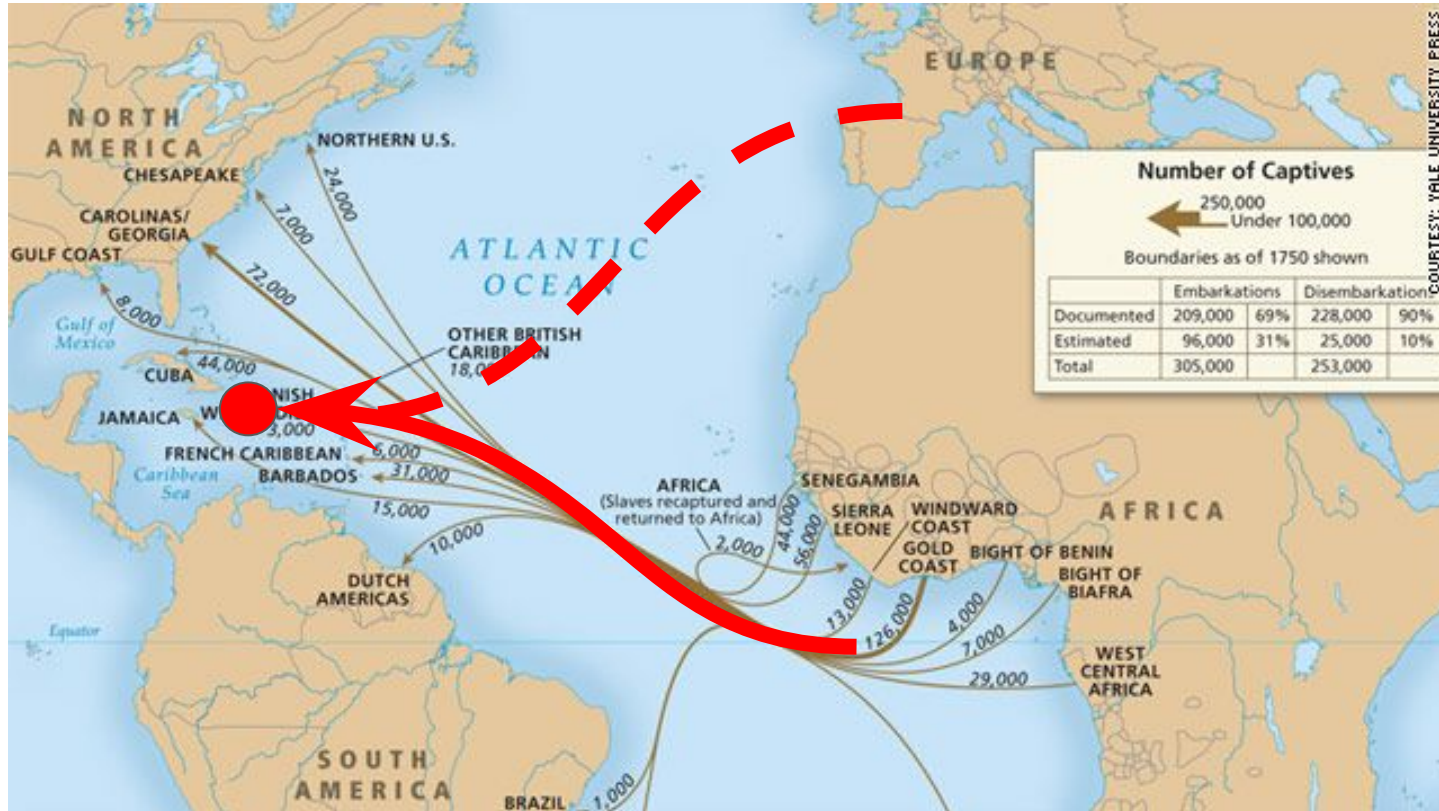**Heather Lent**, Emanuelle Bugliarello, Anders Søgaard

UNIVERSITY OF COPENHAGEN

What is a Creole language?

- Many Creoles share a common, and tragic, history …

What is a Creole language?

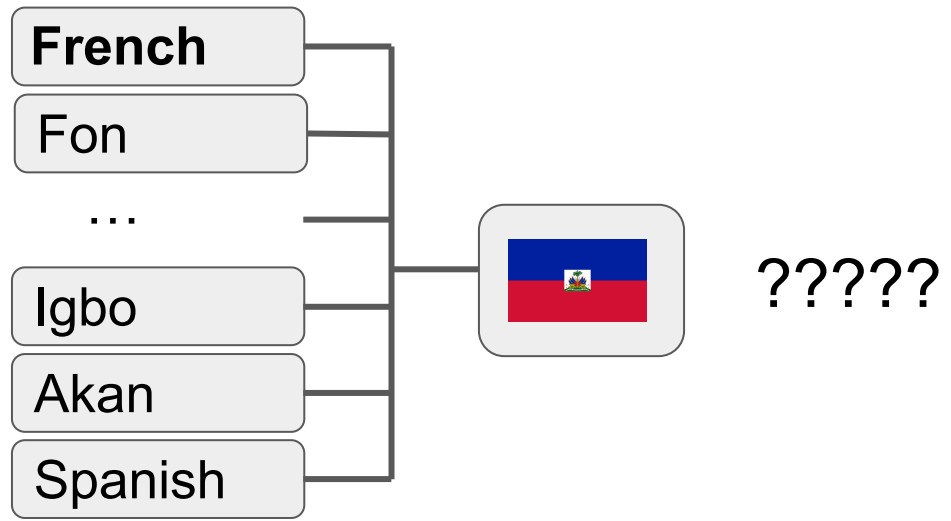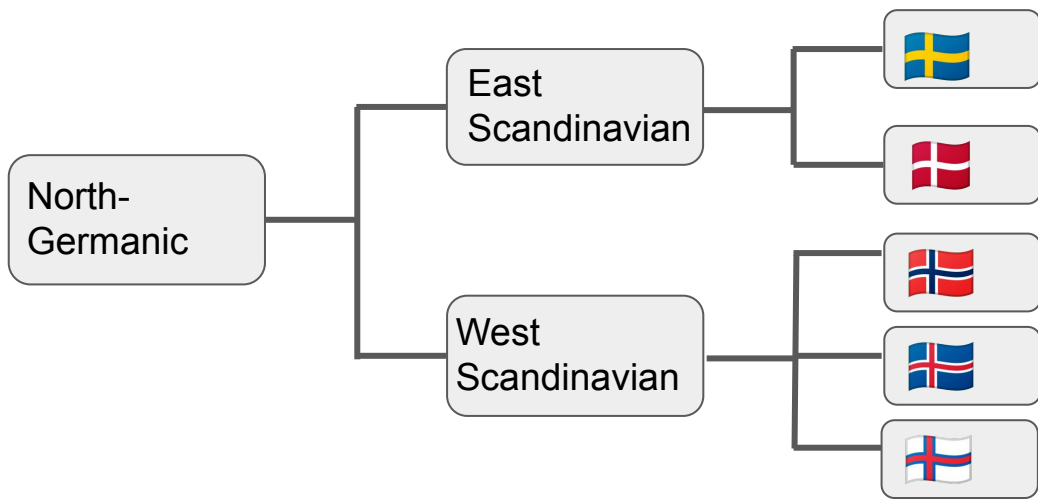- Many Creoles share a common, and tragic, history …

```
North-Germanic
├── East Scandinavian
│   ├── 🇸🇪
│   └── 🇩🇰
└── West Scandinavian
    ├── 🇳🇴
    ├── 🇮🇸
    └── 🇫🇴
```

Non-Creole Phylogeny

?????

R1: Ancestry Transfer Hypothesis

Can we transfer from Ancestor languages to Creoles, similar to how we can facilitate languages between related languages?

Data

- Creoles are low-resource languages
- Bible Corpus (Mayer & Cysouw, 2014)
- Training data:

Type	Target	Training Langs	Train Size (#Sents)
Creole	acf	fra, hau, yor, ibo	38,140
	hat	fra, fon, ibo, spa	31,669
	jam	eng, hau, spa, ibo	44,545
	pcm	eng, hau, yor, por	35,189
Non-Creole	dan	nno, isl, swe, deu	39,354
	spa	fra, por, ita, rom	30,870

- Dev of 500 verses

Data

- Creoles are low-resource languages
- Bible Corpus (Mayer & Cysouw, 2014)
- Training data:

Type	Target	Training Langs	Train Size (#Sents)
Creole	acf	fra, hau, yor, ibo	38,140
	hat	fra, fon, ibo, spa	31,669
	jam	eng, hau, spa, ibo	44,545
	pcm	eng, hau, yor, por	35,189
Non-Creole	dan	nno, isl, swe, deu	39,354
	spa	fra, por, ita, rom	30,870
Control	-	afr, chr, hun, quy	37,398

- Dev of 500 verses

Methods

- A tiny BERT for tiny data (Dufter & Schütz, 2020)
  - 1 Attention head (K et al. 2020)
  - Smaller transformers for smaller datasets (Susanto et al., 2019)

- BERT Word Piece Tokenizer
  - 10,2480 tokens
  - We also experimented with:
    - Size={1024, 2048}
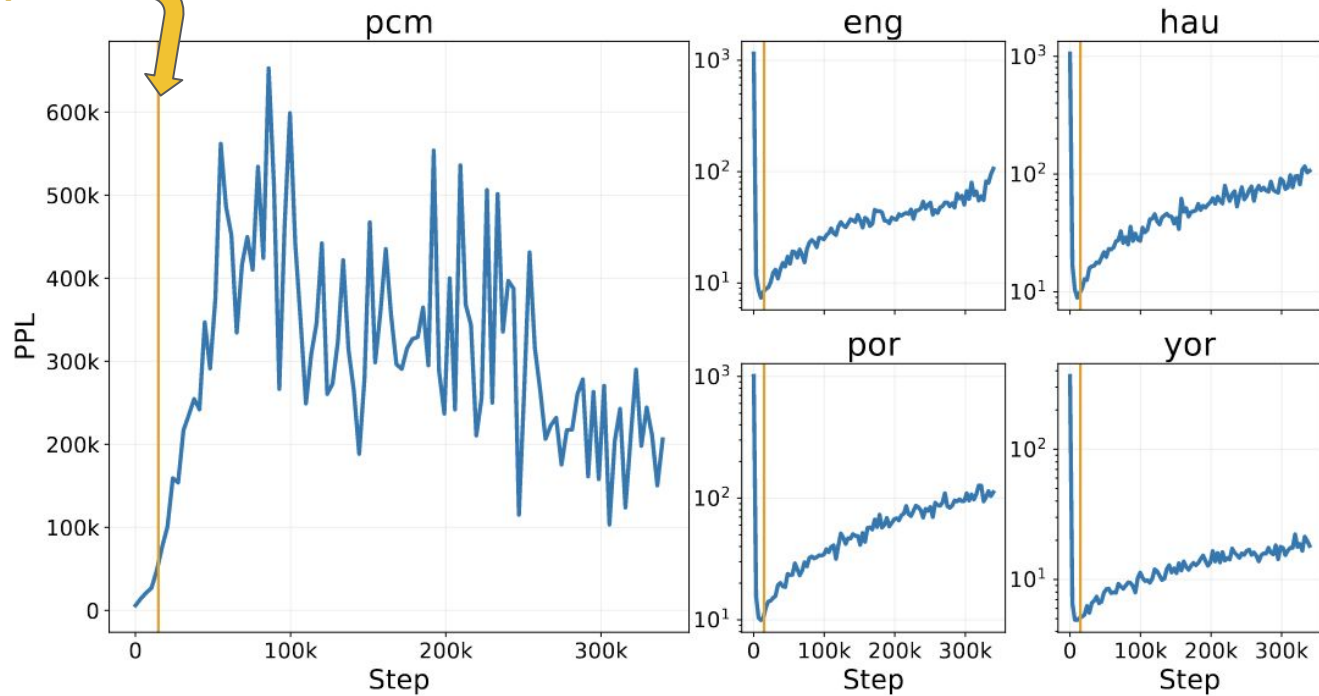   - Segmentation={grapheme-to-phoneme, BPE}

- A note on segmentation …

Methods

- When we train for 100 epochs …

- R1 not easily validated

- What about training for long? Can we benefit from

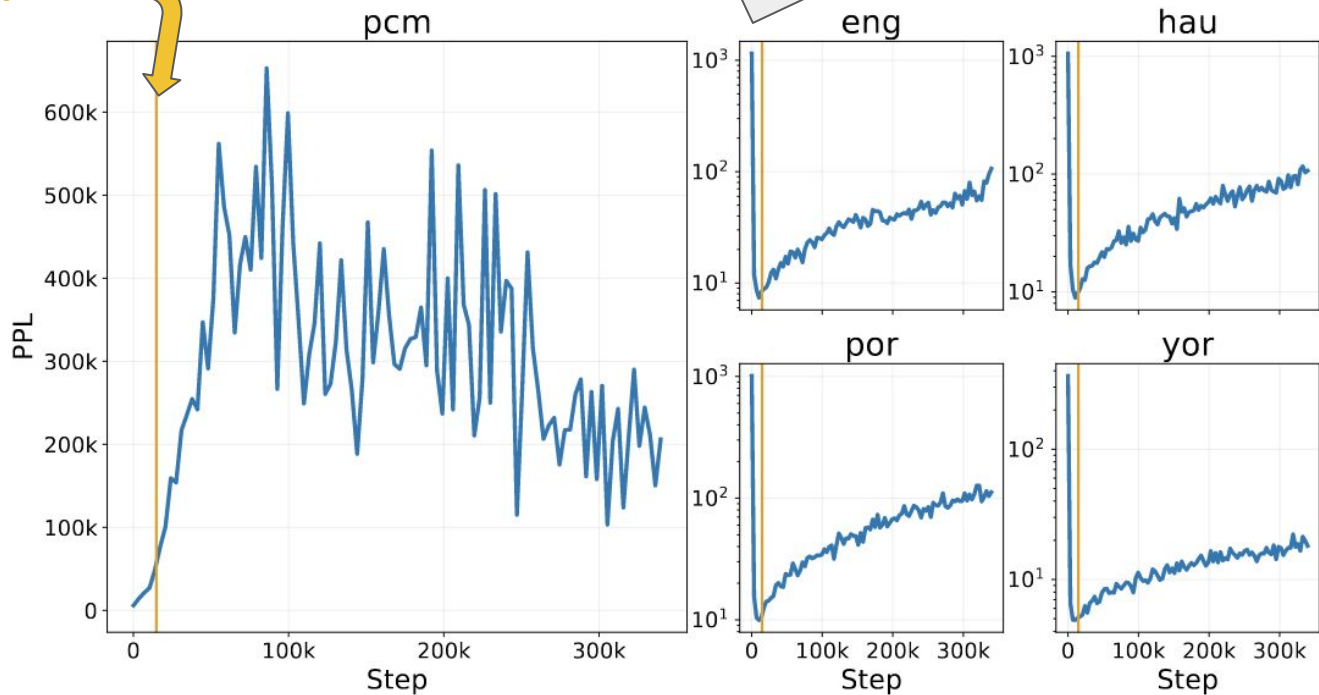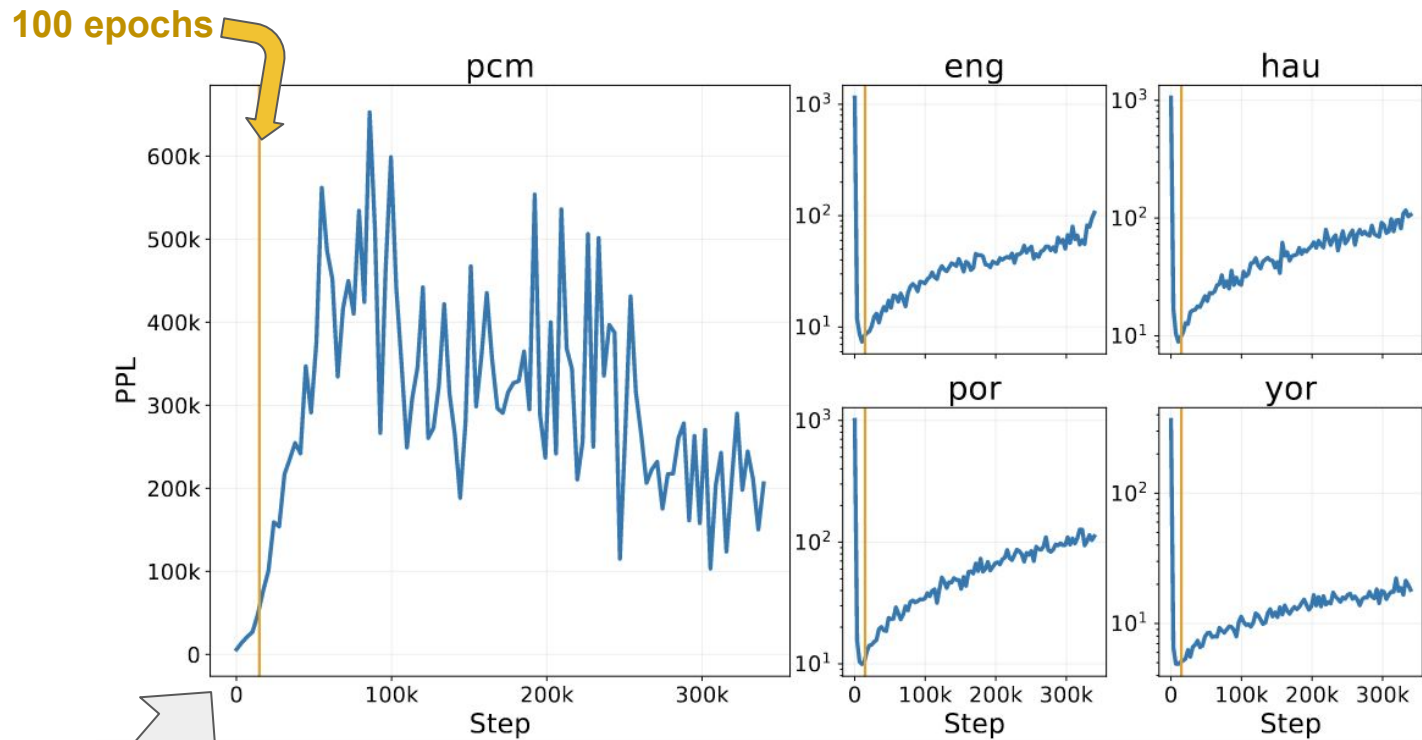- Talk about PPL as a measurement

Results: Training for long

100 epochs

Results: Training for long
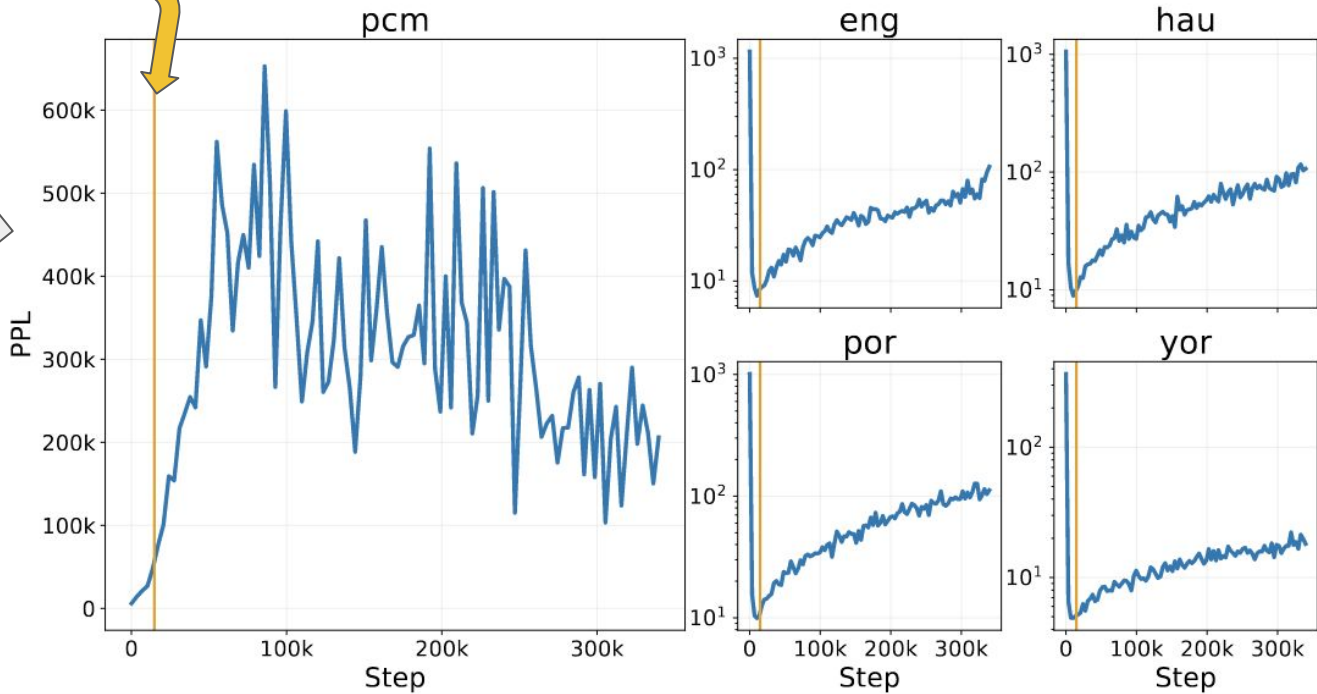
Model learns src langs in 100 epochs

100 epochs

Results: Training for long

100 epochs

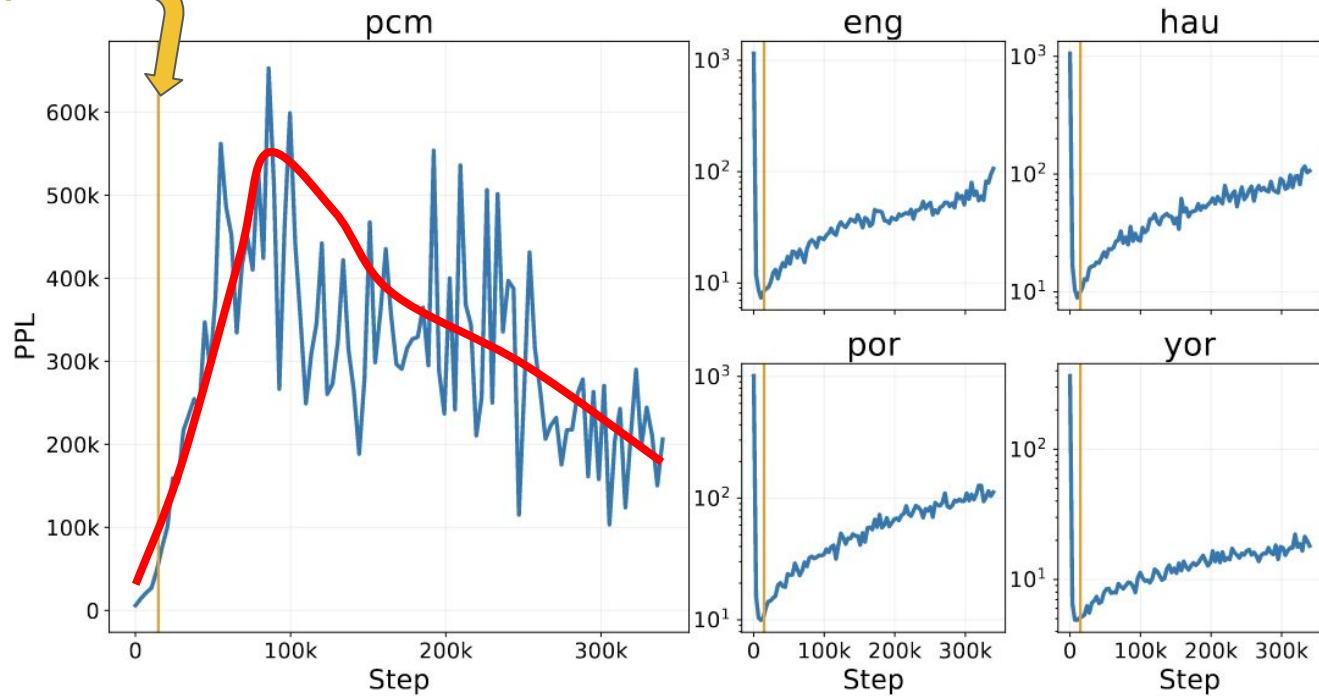Model unable to transfer to Creole in 100 epochs
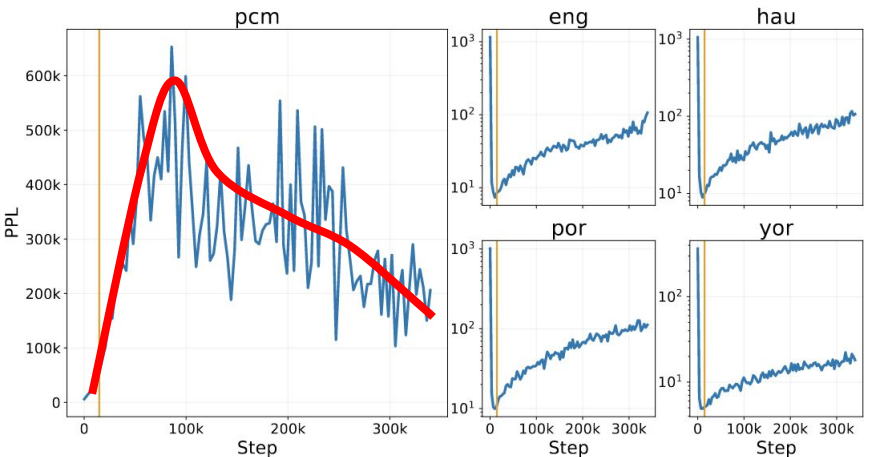
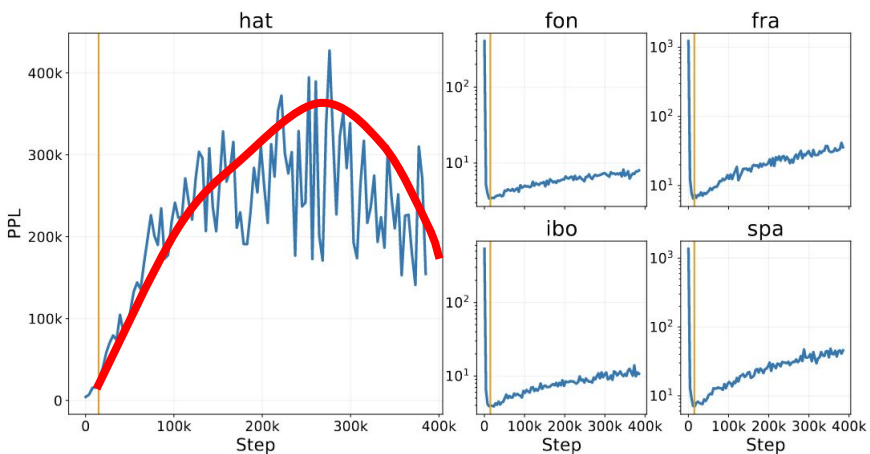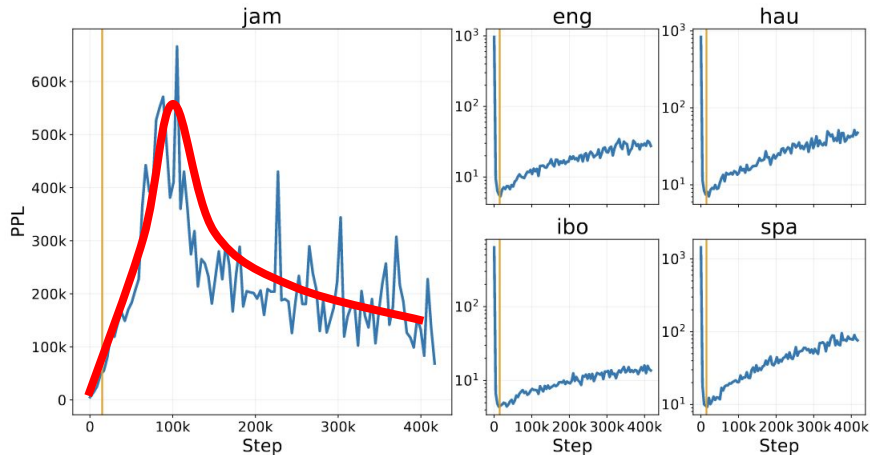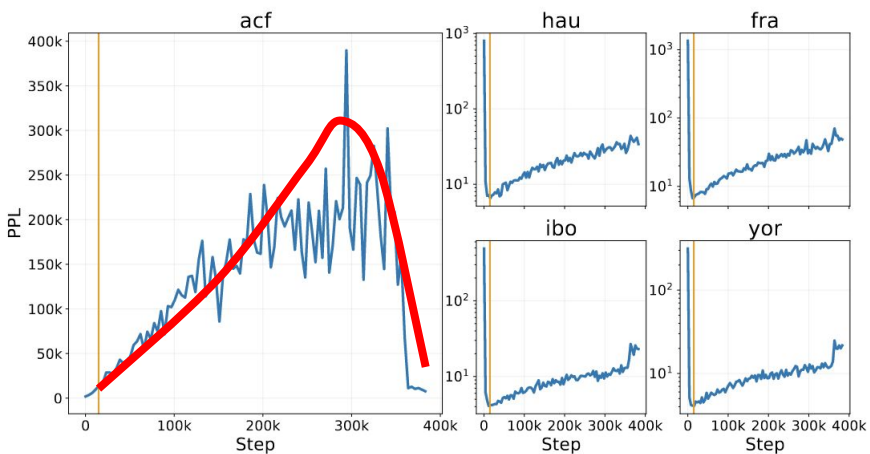# Results: Training for long

100 epochs

Overfitting on Ancestor data leads to very high perplexity

Results: Training for long

100 epochs

Results across Creoles

R2: Ancestry Bottleneck Hypothesis
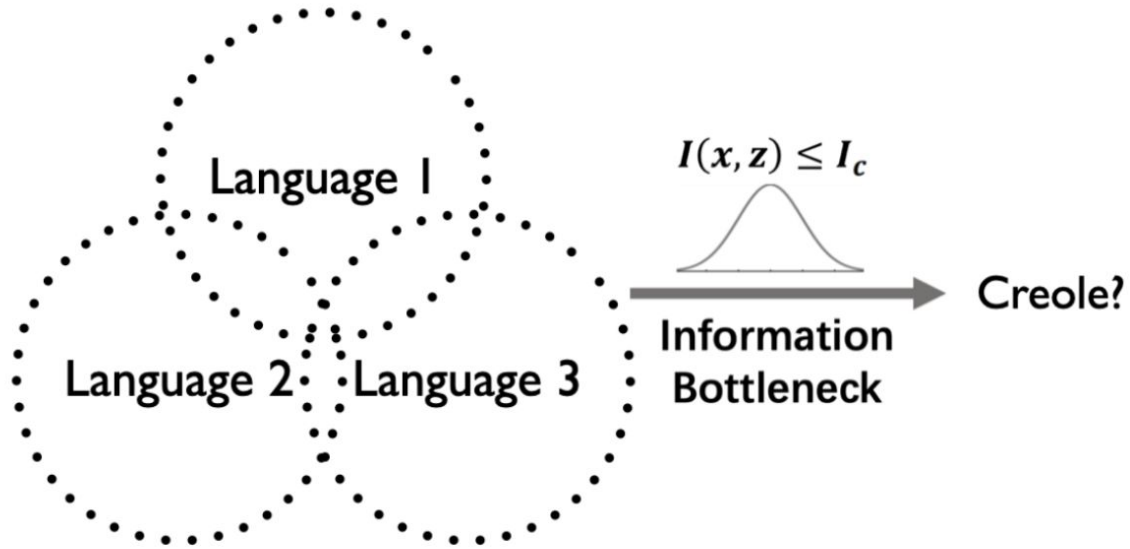
Could this compression be useful?

No, the compression is not useful.

- For downstream tasks [NER on pcm], the "compression"/drop is not helpful. Late stage checkpoints didn't perform better than earlier stage ones.

- And it didn't perform better than pre-trained MBERT (step 0).

- Continued few-shot learning for LM's yields a similar picture
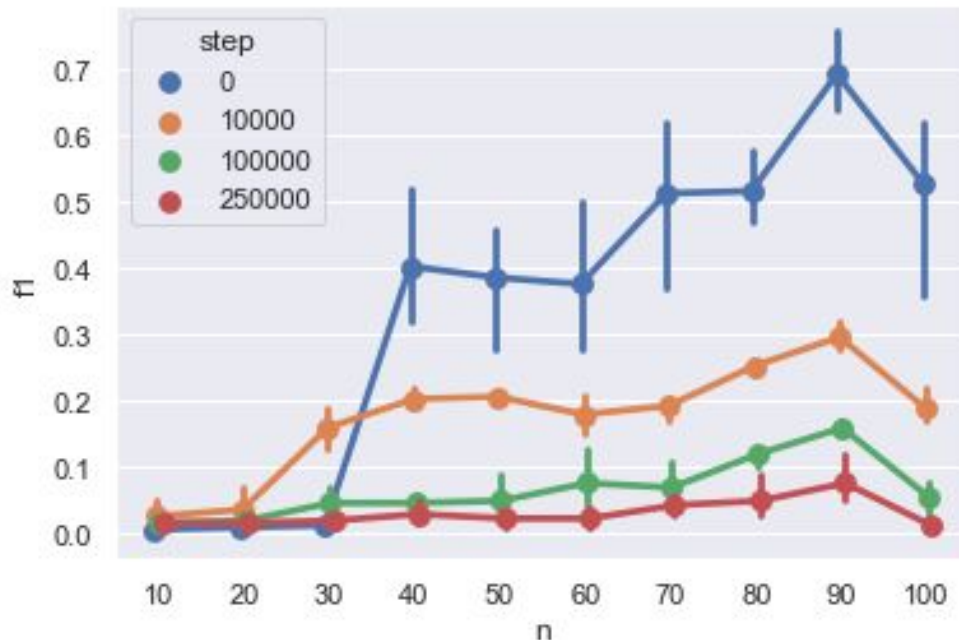
But why might we observe this compression at all?

Creole Similarity: Training on Ancestor vs Random

Target:
Nigerian Pidgin English (pcm)
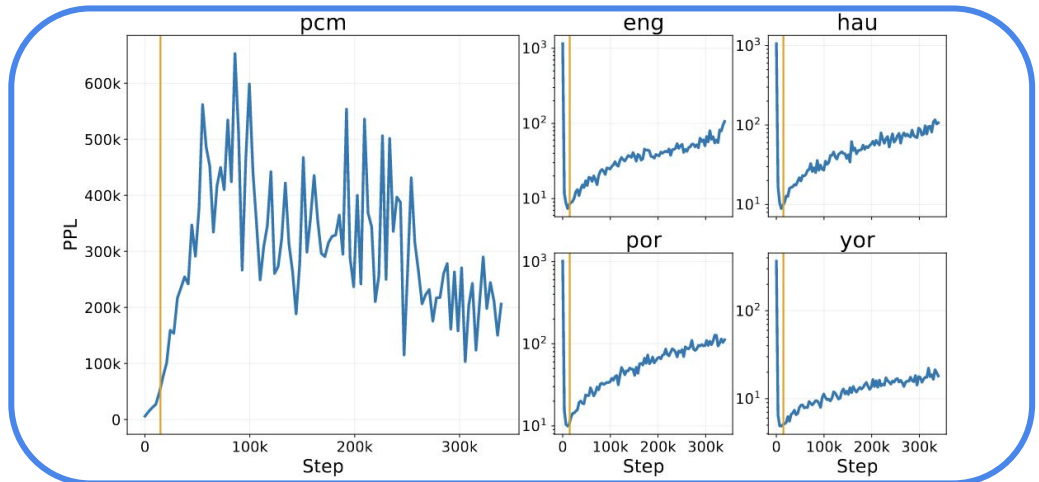
Source:
English, Hausa, Portuguese, Yoruba

Source:
Afrikaans, Cherokee, Hungarian, Quechua

Creole Similarity

WALS Cosine Distance - Related

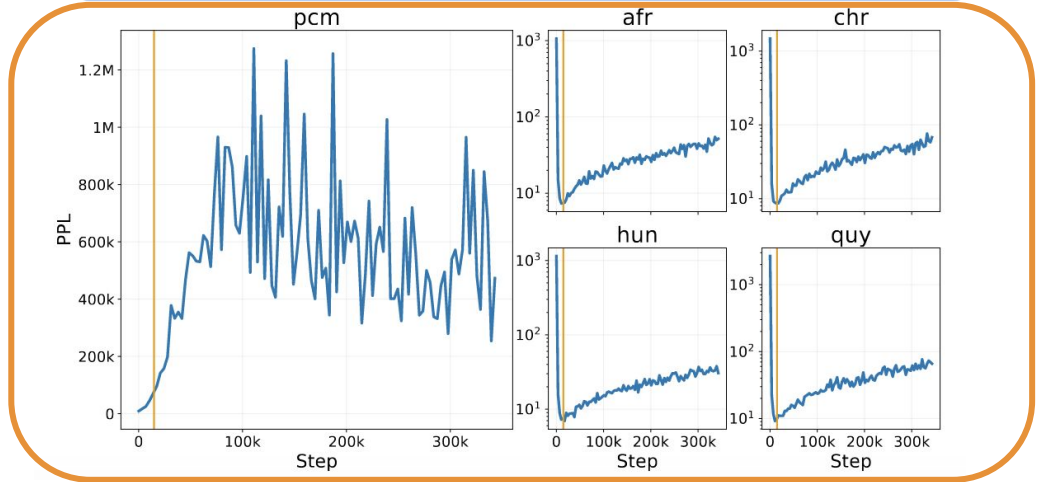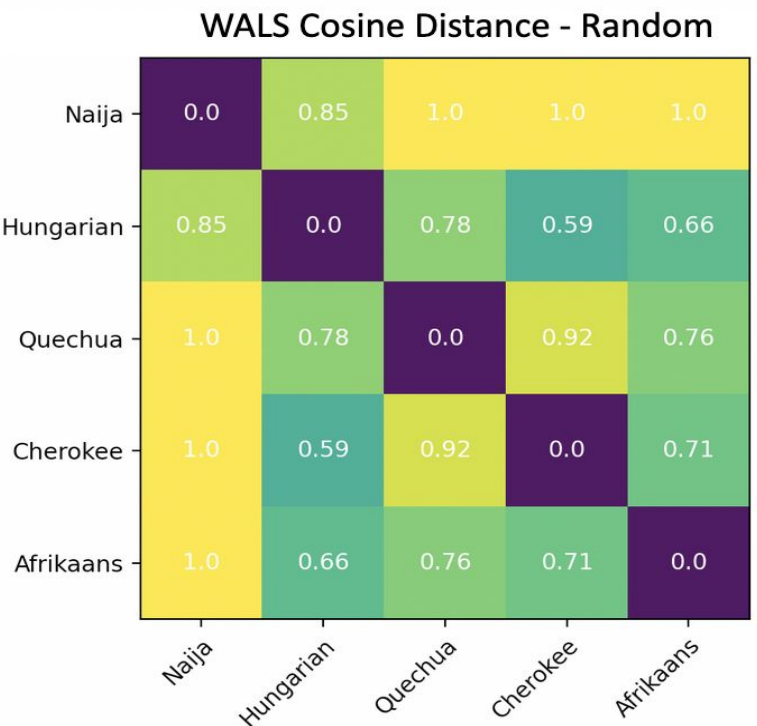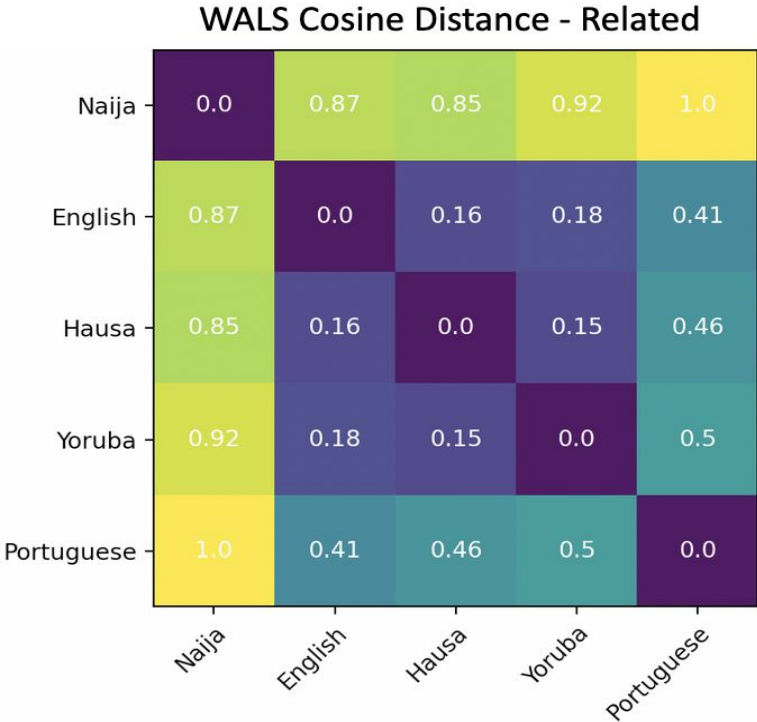|	Naija	English	Hausa	Yoruba	Portuguese
Naija	0.0	0.87	0.85	0.92	1.0
English	0.87	0.0	0.16	0.18	0.41
Hausa	0.85	0.16	0.0	0.15	0.46
Yoruba	0.92	0.18	0.15	0.0	0.5
Portuguese	1.0	0.41	0.46	0.5	0.0

WALS Cosine Distance - Random

	Naija	Hungarian	Quechua	Cherokee	Afrikaans
Naija	0.0	0.85	1.0	1.0	1.0
Hungarian	0.85	0.0	0.78	0.59	0.66
Quechua	1.0	0.78	0.0	0.92	0.76
Cherokee	1.0	0.59	0.92	0.0	0.71
Afrikaans	1.0	0.66	0.76	0.71	0.0

[!!!!] Caveat! WALS is probably a poor resource for this!!!

Take-Away Points

1) Typical cross-lingual transfer methods **do not facilitate transfer** from ancestor languages to Creoles.

2) Creoles can be useful for investigating cross-lingual transfer
(e.g. segmentation)

Creoles

X-lingual learning