On Language Models For Creoles

Heather Lent  
hcl@di.ku.dk

Emanuele Bugliarello  
emanuele@di.ku.dk

Miryam de Lhoneux  
ml@di.ku.dk

Chen Qiu  
chen@wust.edu.cn

Anders Søgaard  
soegaard@di.ku.dk
What are creoles?

- A language born from many parents…
- More than code-switching or mixed language
- Not any less of a language!
What are creoles?

- A language born from many parents…
- More than code-switching or mixed language
- Not any less of a language!

(Often) tragic historical events
What are creoles?

- Examples of creole languages:
  - Nigerian Pidgin English (“Naija”)
  - Singaporean Colloquial English (“Singlish”)
  - Haitian Creole
What are creoles?

- Examples of creole languages:
  - Nigerian Pidgin English (“Naija”)
  - Singaporean Colloquial English (“Singlish”)
  - Haitian Creole

```
| Tamil | Mandarin (我们) | Cantonese (拍拖) | English   | Malay     | Eng | Malay     | Hokkien/Hakka (店) | X |
|-------|---------------|------------------|-----------|-----------|-----|-----------|---------------------|---|
| Dey   | wǒ men        | paktor           | always    | makan     | at  | kopi      | one                 |<INTJ>|
| Hey   | we            | date             | always    | eat       | at  | coffee    | shop                |   |

Standard English: “Hey, when we date we always eat at the coffee shop”
What are creoles?

- Examples of creole languages:
  - Nigerian Pidgin English ("Naija")
  - Singaporean Colloquial English ("Singlish")
  - Haitian Creole

Bajpai et al. (2017):

(1) John sibei hum sup one.
(2) John very buaya sia.
Why work on creoles?

● Interesting for cross-lingual and multilingual NLP
  ○ Relationships/dynamics between parent languages and creole
  ○ Creole continuum (basillect, mesolect, acrolect)

● Expanding NLP
  ○ Low resource languages
  ○ Often linguae francae
  ○ Challenging idea that creoles are “degenerate” (low prestige)

● Other reasons
  ○ NLP for crisis management
Also the fact that hundreds of millions of people speak creole languages around the world!

Image from “The Atlas of Pidgin and Creole Languages Structure Online” at apics-online.info
Creoles, Demographics, and DRO

- Creoles made from collection of different languages

- Some languages more dominant than others (e.g. lexifier), but other languages still contribute to a creole’s vocabulary, syntax, etc.

- In Algorithmic Fairness, “Distributionally Robust Optimization” (DRO) aims to protect minority groups by minimizing loss on each group, rather than averaging across all data.

- “Distributionally Robust Language Modeling” by Oren et al. 2019.
Can DRO help us create better LMs for creoles, which are more robust to the language dynamics at hand?
## DRO for Creole

| DRO-Language | DRO-One* | DRO-Random* |
|--------------|----------|-------------|
| Language id  | All examples in one group | Assign examples a random group ID |
### DRO for Creole

|                | DRO-Language | DRO-One* | DRO-Random* |
|----------------|--------------|----------|-------------|
| Mixed-Languages| Language id  | All examples in one group | Assign examples a random group ID |
| Creole-Only    |              | “        | “           |
DRO for Creole

| Mixed-Languages | DRO-Language | DRO-One* | DRO-Random* |
|-----------------|--------------|----------|-------------|
| Language id     | All examples in one group | Assign examples a random group ID |

| Creole-Only       | Fasttext language identification |
|-------------------|---------------------------------|
| “Pikin wey like to play wit wetin no dey common and sabi one particular subject reach ground” | “” |

- en: 87.46%
- pt: 0.23%
- yo: 0.03%
| BERT          | Nigerian Pidgin | Singlish | Haitian Creole |
|--------------|-----------------|----------|----------------|
|              | P@1  | P_D@1 | PLL     | P@1   | P_D@1 | PLL     | P@1   | P_D@1 | PLL     |
| Pretrained   | 22.79 | 10.92 | 142.65  | 23.94 | 21.09 | 76.01   | 18.84 | 5.65  | 177.40  |
| **ERM**      | **63.83** | **59.97** | **42.41** | **46.77** | **42.89** | **41.06** | **68.09** | **43.35** | **55.04** |
| DRO-One      | 60.99 | 56.76 | 52.51   | 44.23 | 40.73 | 49.18   | 57.04 | 36.73 | 121.51  |
| DRO-Random   | 60.40 | 56.33 | 52.69   | 43.33 | 39.07 | 49.14   | 57.65 | 36.16 | 119.17  |
| DRO-Language | 60.40 | 54.80 | 54.17   | 43.19 | 39.57 | 48.88   | 57.55 | 36.69 | 118.85  |
| **ERM**      | **73.72** | **71.38** | **28.14** | **53.80** | **51.26** | **34.22** | **73.15** | **55.50** | **55.51** |
| DRO-One      | 64.28 | 59.86 | 61.81   | 45.34 | 43.59 | 66.53   | 58.16 | 36.91 | 144.46  |
| DRO-Random   | 63.72 | 59.31 | 60.31   | 45.73 | 42.40 | 64.16   | 57.65 | 37.41 | 142.04  |
| DRO-Language | 63.58 | 59.74 | 56.82   | 44.73 | 40.57 | 53.72   | 56.94 | 35.50 | 138.60  |

Table 3: Intrinsic evaluation: Precision@1 (P@1), Precision@1 for words in our creole dictionary (P_D@1), and average Pseudo-log-likelihood score (PLL). We report results for MIXED-LANGUAGE (top) and CREOLE-ONLY (bottom). We note that ERM consistently outperforms the language models trained with robust objectives.
## Results (Extrinsic Evaluation)

| BERT       | Nigerian Pidgin | Singlish |
|------------|-----------------|----------|
|            | NER [F₁] | UPOS [Acc] | UPOS [Acc] |
| MIXED      |            |           |          |
| ERM        | 87.86     | 98.00     | 91.24    |
| DRO-Language | **88.40** | **98.06** | 90.22    |
| C-ONLY     |            |           |          |
| ERM        | 87.98     | 98.04     | 91.17    |
| DRO-Language | 87.12  | 97.98     | 90.44    |

Table 4: Extrinsic evaluation. Similar performance on downstream tasks across all models demonstrate show that language model training did not benefit significantly from neither DRO nor data in related languages.
When DRO fails...

● Overparameterization?
● Regularization?
● Creole instability and domain drift?
When DRO fails...

- Overparameterization?
- Regularization?
- Creole instability and domain drift?

| Language       | Domain-1                     | Domain-2 | PAD  |
|----------------|------------------------------|----------|------|
| English        | Disaster Response Corpus     | Newswire| 1.75 |
| Haitian Creole | Disaster Response Corpus     | Newswire| 1.47 |
| English        | EWT-UD                       | NUD      | 1.04 |
| Nigerian       | UNMT                         | NUD      | 1.28 |
Discussion & Conclusions

- Our results show that vanilla ERM is better than DRO for LM of creoles
- Likely the result of the relative stability of creole languages
- There is much interesting work to be done for creole NLP! Especially w.r.t. modeling dynamics specific to creoles (e.g. development, social factors, etc.), and especially in cross-lingual and multilingual NLP
- Hope we have inspired you to work on creoles :-)

THANK YOU
Extra Slides
(not part of presentation)
Creoles, Demographics, and DRO

- Creoles made from collection of different languages
- Some languages more dominant than others (e.g., lexifier), but other languages still contribute to a creole's vocabulary, syntax, etc.
- In Algorithmic Fairness, "Distributionally Robust Optimization" (DRO) aims to protect minority groups by minimizing loss on each group, rather than averaging across all data.
- "Distributionally Robust Language Modeling" by Oren et al. 2019

![Figure 1. Illustration of a training corpus as a density (black) with mostly news stories (red) and a small number of restaurant reviews (blue). The standard MLE model (gray) reflects the underlying data and assigns little weight to reviews, and thus performs poorly on reviews. A more robust model should try to equalize the weight across all topics so that it can perform well regardless of which topics appear at test time.](image)

Figure borrowed from “Distributionally Robust Language Modeling” By Oren et al. 2019

Yonatan Oren, Shiori Sagawa, Tatsunori B. Hashimoto, and Percy Liang. 2019. EMNLP.
## Data

| Language      | Source                  | Domain |
|---------------|-------------------------|--------|
| en, fr, es, pt, yo, zh, ta | WMT-News 2020           | news   |
| ms            | Malay 30k News          | news   |
| Nigerian Pidgin | PidginUNMT Corpus     | news   |
| Singlish      | Singapore SMS Corpus    | sms    |
| Haitian Creole| Disaster Response Corpus| sms    |

Table 1: Data resources utilized in our experiments.

| Creole          | Langs | # Train Mixed-Lang | # Train Creole-Only | # Dev Creole-Only |
|-----------------|-------|--------------------|---------------------|-------------------|
| Nigerian Pidgin | en, pt, yo | 230,105            | 53,006              | 3,359             |
| Singlish        | en, zh, ms, ta | 265,030            | 67,615              | 2,790             |
| Haitian Creole  | fr, yo, es    | 32,768             | 8,192               | 988               |

Table 2: Creoles, their influential languages (Langs), and the number of examples in the Train-Dev split for our **Mixed-Language** and **Creole-Only** experiments. Both use the same creole-only dev dataset.
Figure 3: Distributions of identified languages across the CREOLE-ONLY test set. **Top:** distributions for the influential languages included in MIXED-LANGUAGE. **Bottom:** distributions of the five languages that had the highest prediction scores for each creole, where we see a bias towards European languages.
### Results (Intrinsic Evaluation)

|    | Nigerian Pidgin | Singlish | Haitian Creole |
|----|-----------------|----------|---------------|
|    | P@1  | P_D@1 | PLL | P@1  | P_D@1 | PLL | P@1  | P_D@1 | PLL |
| Pretrained | 22.79 | 10.92 | 142.65 | 23.94 | 21.09 | 76.01 | 18.84 | 5.65 | 177.40 |
| **ERM** | **63.83** | **59.97** | **42.41** | **46.77** | **42.89** | **41.06** | **68.09** | **43.35** | **55.04** |
| DRO-One | 60.99 | 56.76 | 52.51 | 44.23 | 40.73 | 49.18 | 57.04 | 36.73 | 121.51 |
| DRO-Random | 60.40 | 56.33 | 52.69 | 43.33 | 39.07 | 49.14 | 57.65 | 36.16 | 119.17 |
| DRO-Language | 60.40 | 54.80 | 54.17 | 43.19 | 39.57 | 48.88 | 57.55 | 36.69 | 118.85 |
| **ERM** | **73.72** | **71.38** | **28.14** | **53.80** | **51.26** | **34.22** | **73.15** | **55.50** | **55.51** |
| DRO-One | 64.28 | 59.86 | 61.81 | 45.34 | 43.59 | 66.53 | 58.16 | 36.91 | 144.46 |
| DRO-Random | 63.72 | 59.31 | 60.31 | 45.73 | 42.40 | 64.16 | 57.65 | 37.41 | 142.04 |
| DRO-Language | 63.58 | 59.74 | 56.82 | 44.73 | 40.57 | 53.72 | 56.94 | 35.50 | 138.60 |

Table 3: Intrinsic evaluation: Precision@1 (P@1), Precision@1 for words in our creole dictionary (P_D@1), and average Pseudo-log-likelihood score (PLL). We report results for **MIXED-LANGUAGE** (top) and **CREOLE-ONLY** (bottom). We note that ERM consistently outperforms the language models trained with robust objectives.
|                | Nigerian Pidgin |                  | Singlish |                  | Haitian Creole |                  |
|----------------|-----------------|-----------------|----------|-----------------|---------------|-----------------|
|                | P@1  | P_D@1 | PLL    | P@1  | P_D@1 | PLL    | P@1  | P_D@1 | PLL    |
| Pretrained     | 22.79 | 10.92 | 142.65 | 23.94 | 21.09 | 76.01  | 18.84 | 5.65  | 177.40 |
| ERM            | 63.83 | 59.97 | 42.41  | 46.77 | 42.89 | 41.06  | 68.09 | 43.35 | 55.04  |
| DRO-One        | 60.99 | 56.76 | 52.51  | 44.23 | 40.73 | 49.18  | 57.04 | 36.73 | 121.51 |
| DRO-Random     | 60.40 | 56.33 | 52.69  | 43.33 | 39.07 | 49.14  | 57.65 | 36.16 | 119.17 |
| DRO-Language   | 60.40 | 54.80 | 54.17  | 43.19 | 39.57 | 48.88  | 57.55 | 36.69 | 118.85 |
| ERM            | 73.72 | 71.38 | 28.14  | 53.80 | 51.26 | 34.22  | 73.15 | 55.50 | 55.51  |
| DRO-One        | 64.28 | 59.86 | 61.81  | 45.34 | 43.59 | 66.53  | 58.16 | 36.91 | 144.46 |
| DRO-Random     | 63.72 | 59.31 | 60.31  | 45.73 | 42.40 | 64.16  | 57.65 | 37.41 | 142.04 |
| DRO-Language   | 63.58 | 59.74 | 56.82  | 44.73 | 40.57 | 53.72  | 56.94 | 35.50 | 138.60 |

Table 3: Intrinsic evaluation: Precision@1 (P@1), Precision@1 for words in our creole dictionary (P_D@1), and average Pseudo-log-likelihood score (PLL). We report results for MIXED-LANGUAGE (top) and CREOLE-ONLY (bottom). We note that ERM consistently outperforms the language models trained with robust objectives.
## Results

| BERT       | Nigerian Pidgin | Singlish | Haitian Creole |
|------------|-----------------|----------|----------------|
|            | P@1  | P_{D}@1 | PLL     | P@1  | P_{D}@1 | PLL     | P@1  | P_{D}@1 | PLL     |
| Pretrained | 22.79 | 10.92   | 142.65  | 23.94 | 21.09   | 76.01   | 18.84 | 5.65    | 177.40  |
| **ERM**    | 63.83 | 59.97   | 42.41   | 46.77 | 42.89   | 41.06   | 68.09 | 43.35   | 55.04   |
| DRO-One    | 60.99 | 56.76   | 52.51   | 44.23 | 40.73   | 49.18   | 57.04 | 36.73   | 121.51  |
| DRO-Random | 60.40 | 56.33   | 52.69   | 43.33 | 39.07   | 49.14   | 57.65 | 36.16   | 119.17  |
| DRO-Language | 60.40 | 54.80   | 54.17   | 43.19 | 39.57   | 48.88   | 57.55 | 36.69   | 118.85  |
| **ERM**    | 73.72 | 71.38   | 28.14   | 53.80 | 51.26   | 34.22   | 73.15 | 55.50   | 55.51   |
| DRO-One    | 64.28 | 59.86   | 61.81   | 45.34 | 43.59   | 66.53   | 58.16 | 36.91   | 144.46  |
| DRO-Random | 63.72 | 59.31   | 60.31   | 45.73 | 42.40   | 64.16   | 57.65 | 37.41   | 142.04  |
| DRO-Language | 63.58 | 59.74   | 56.82   | 44.73 | 40.57   | 53.72   | 56.94 | 35.50   | 138.60  |

Table 3: Intrinsic evaluation: Precision@1 (P@1), Precision@1 for words in our creole dictionary (P_{D}@1), and average Pseudo-log-likelihood score (PLL). We report results for MIXED-Language (top) and CREOLE-ONLY (bottom). We note that ERM consistently outperforms the language models trained with robust objectives.
Overparameterization

| BERT        | Size | Nigerian Pidgin |       |       |       |
|-------------|------|-----------------|-------|-------|-------|
|             |      | P@1            | P_D@1| PLL   |
| ERM         | Tiny | 31.31           | 26.12 | 110.23|
|             | Small| 47.39           | 46.75 | 77.47 |
|             | Base | 63.83           | 59.97 | 42.41 |
| DRO-Language| Tiny | 31.00           | 23.09 | 99.70 |
|             | Small| 43.00           | 37.75 | 82.50 |
|             | Base | 60.40           | 54.80 | 54.17 |

Table 5: Over-parameterization experiments with **MIXED-LANGUAGE** Nigerian Pidgin English data. Smaller sized models do not benefit DRO over ERM.
Regularization

| BERT          | Weight Decay | Nigerian Pidgin |
|---------------|--------------|-----------------|
|               |              | P@1  | P_D@1 | PLL  |
| ERM           | 0.01         | 47.39 | 46.75 | 77.47|
|               | 0.01         | 43.00 | 37.75 | 82.50|
| DRO-Language  | 0.05         | 42.86 | 38.47 | 83.03|
|               | 0.10         | 43.00 | 38.74 | 81.80|
|               | 0.30         | 42.70 | 39.53 | 81.94|

Table 6: Regularization experiments on MIXED-LANGUAGE Nigerian Pidgin data, based on BERT_{Small}. 


| Language       | Domain-1                     | Domain-2 | PAD  |
|---------------|------------------------------|----------|------|
| English       | Disaster Response Corpus     | Newswire| 1.75 |
| Haitian Creole| Disaster Response Corpus     | Newswire| 1.47 |
| English       | EWT-UD                       | NUD      | 1.04 |
| Nigerian      | UNMT                         | NUD      | 1.28 |

Table 7: Proxy $A$-distance (PAD) scores on parallel (Haitian) or near-parallel (Nigerian) data. PAD is proportional to domain classification error; hence, large distances mean high domain divergence. Our results suggest that creole languages do *not* exhibit significantly more drift than other languages.