

How to Parse a Creole: When Martinican Creole Meets French

Ludovic Mompelat
Indiana University
lmompela@iu.edu

Daniel Dakota
Indiana University
ddakota@iu.edu

Sandra Kübler
Indiana University
skuebler@indiana.edu

Abstract

We investigate methods to develop a parser for Martinican Creole, a highly under-resourced language, using a French treebank. We compare transfer learning and multi-task learning models and examine different input features and strategies to handle the massive size imbalance between the treebanks. Surprisingly, we find that a simple concatenated (French + Martinican Creole) baseline yields optimal results even though it has access to only 80 Martinican Creole sentences. POS embeddings work better than lexical ones, but they suffer from negative transfer.

1 Introduction

Syntactic analysis is an essential task for language documentation and language revitalization, as it allows a deeper understanding of languages. Under-resourced languages often suffer from the lack of annotated gold standard data available to develop and offer NLP solutions for the communities speaking the language. Moreover, there is a low number of researchers trained in formal linguistics and/or linguistic annotations which causes additional challenges in the creation of language resources for such languages. In recent years, research on parsing has developed a focus on under-resourced languages (Agić et al., 2016; Vania et al., 2019; Meechan-Maddon and Nivre, 2019), but creoles have received less attention.

In this study, we develop a dependency parsing model for Martinican Creole (MC), a French-based Creole, mostly spoken in Martinique, a French island in the Caribbean. Being a French territory, French and MC coexist in an unbalanced manner. The diglossic situation makes French the dominant language in many contexts, although in the past decades Martinican Creole has seen an expansion of its communicative contexts (Bernabé, 2004; Véronique, 2020). This is due to the codification and standardization processes the language

underwent, especially by the [GEREC \(1982\)](#), and to the linguistic policies developed in an effort to safeguard and revitalize this language. However, one aspect of this revitalization process that is currently missing is the expansion of NLP tools and resources for MC (and other creole languages). Creole languages based on the same lexifier language (in our case French) are extremely similar and in many cases, mutually intelligible. Thus, developing NLP solutions for one creole language provides a basis to transfer knowledge to other related Creole languages.

The goal of this project is to investigate the best methods for developing a parser for an extremely low resource language when this language is a creole language. In our case, the creole is Martinican Creole. The main question here is whether the lexifier language, i.e., French, is similar enough to serve as basis for training a parser without further modification.

2 Research Questions

Our overarching research question is the following: Can we leverage a French treebank using transfer learning or multi-task learning approaches to create a parser model for Martinican Creole given that we only have a very small treebank? Can we leverage the similarity between the creole and its lexifier language, French?

To answer this question, we need to answer the following questions:

1. Which types of embeddings can be used? Given the differences in spelling, are character embeddings, POS embeddings, or BERT embeddings the best representation of the input sentence? (We will not consider multilingual BERT models since the closest language is French, and we have access to large French embeddings models.)
2. In a transfer learning setting, how do we best

use the very limited Martinican Creole data? Is it worth the effort to annotate data for optimizing the parser, or can we optimize it on French? Is there enough structural and lexical similarity between French and the creole to make this possible?

3. In a transfer learning setting, how do we deal with the extreme imbalance between the large French Treebank and the small Martinican Creole Treebank? Can we prevent the parser from overfitting?
4. Can we leverage a multi-task learning model to handle the imbalance between French and the creole? More specifically, will loss weighting be able to counterbalance the treebank sizes?
5. Can we determine the linguistic characteristics of Martinican Creole that provide challenges to parsers based on standard transfer learning and on multi-task learning?

3 Related Work

Creoles are still under-researched in NLP. Noticeable work includes language model comparisons by [Lent et al. \(2021\)](#) between Haitian Creole, Nigerian Pidgin English, and Singaporean Colloquial English, trained with empirical risk minimization, against language models with distributionally robust ones, finding that the former performed better for Creoles. One reason postulated may be the absence of drift due to the relative stability of creoles.

Regarding French-based creoles, Haitian Creole was the subject of an extensive collaboration in Machine Translation led by Microsoft Research ([Lewis, 2010](#)) following the 2010 earthquake. [Milour and Fort \(2018\)](#) led a project of crowdsourcing of POS tags for Guadeloupean Creole in which they describe the necessary steps and methodology to crowdsource a language for POS tagging. They were able to collect a corpus of nearly 2,500 tokens POS tagged and create a POS tagger reaching 84% accuracy.

The lack of available creole treebanks, with Nigerian Pidgin English ([Caron et al., 2019](#)) the only publicly available Universal Dependency treebank, means that best parsing strategies for Creoles are still being developed. Given the lack of available data, parsing creoles can be viewed as similar to the need to leverage related treebanks

to try and increase performance on the target treebank. A common approach is to concatenate available treebanks and optimize towards the target treebank. This has demonstrated gains in both monolingual ([Björkelund et al., 2017](#); [Velldal et al., 2017](#)) and cross-lingual ([Das et al., 2017](#)) experiments. Another successful technique is to instead train a model on a source treebank and then fine-tune on the target treebank ([Shi et al., 2017](#); [Che et al., 2017](#)).

The most directly related works to ours are [Wang et al. \(2017, 2019\)](#) since they parse Singlish, an English-based Creole, by leveraging its lexifier language, English, to boost performance. [Wang et al. \(2017\)](#) propose a neural stacking architecture which yielded promising results which were further investigated by [Wang et al. \(2019\)](#). They tripled the size of their original Singlish treebank by web scraping and annotating more data and performed additional multi-task experiments for integrating English syntactic knowledge. While multi-task models showed some success, neural stacking methods were still better, as was simply concatenating English and Singlish treebanks in some experiments. Such neural stacking architectures with additional POS information also have helped in the related task of parsing Hindi-English Code-switching data ([Bhat et al., 2018](#)). As far as we know, we are the first to approach the task of dependency parsing a French-based Creole.

4 Properties of Martinican Creole

Martinican Creole (MC) is a French-based creole and part of the Atlantic Creoles language family. Syntactically, MC is an SVO language and is closely related to French, other creoles such as Guadeloupean, Marie Galante, St. Barth, Saint Lucian Creoles, and Haitian Creole, and to a lesser degree to African languages. The differences between MC and the closely related Antillean creoles are mostly lexical, they share very similar syntactic structures.

While MC originates from French, both languages show noticeable syntactic differences, especially wrt. the word order in noun phrases.

- | | | |
|-----|----------------------------|----------|
| (1) | Zanmi-mwen enmen liv-la | (MC) |
| | friends-my like book-the | |
| | Mes amis aime le livre | (French) |
| | "My friends like the book" | |

Example (1) shows a sentence in Martinican Cre-

ole. It demonstrates that determiners like *-la* and modifying pronouns like *-mwen* are post-posed, compared to their French and English counterparts *mes* (my) and *le* (the).

Despite these differences in morpheme order, it is still relatively easy to see the direct parallels between both languages. This makes French a good candidate for a transfer learning approach to parsing MC.

MC is considered a morphologically reduced language (Hazaël-Massieux, 2002): Tense, mood and aspect features are expressed as separate morphemes/markers instead of inflections on the verbal element. There is also no morphological gender/number marking on nouns and adjective.

- (2) Asiparé yo **té ké** vann prop
 Apparently 3PL PST.FUT sell own
 frè-**yo** épi sè-**yo**.
 brothers-3PL and sisters-3PL
 A ce qu'il parait, ils vendraient **leurs** propres frères et **leurs** soeurs.
 "Apparently, they would sell their own siblings"

In example (2), we see that the conditional is expressed in MC by a morpheme combination of the Past/Perfective marker *té* and the Future/Irrealis marker *ké* whereas in French, the conditional is expressed synthetically by the affix *-raient* attached to the end of the verb *vendre*. We also see that general plural nouns like *frè-yo* and *sè-yo* are not morphologically marked in MC, and neither is their accompanying adjective *prop*, whereas in French *frères* and *soeurs* and *propres* are all morphologically marked for gender and number.

Finally, while MC uses a different spelling system from French, the MC pronunciation is much closer to its spelling than in French. MC acquired most of its lexicon from French. Lexical transfer was either phonetically transparent or underwent reanalysis via several phono-lexical processes (such as agglutination (see example (3)), apheresis (see example (4)), syncope (see example (5)), etc.).

- (3) **Agglutination**
 diri [di.ʁi] (MC)
 du riz [dy.ʁi] (French)
 (some) rice
- (4) **Apheresis**
 limen [li.mɛ̃] (MC)
 allumer [a.ly.me] (French)
 to turn on

- (5) **Syncope**
 dòmi [dɔ.mi] (MC)
 dormir [dɔ.ʁmi] (French)
 To sleep

In both cases, while the lexical transfer can easily be identified at the phonetic level, it is a more difficult to identify at the orthographic level, since there are significant differences in the respective spelling systems. Because of the amount of differences, it is possible that French embeddings may not be useful, since there may not be enough lexical overlap between French and MC, even on the subword level.

5 Methodology

5.1 Treebanks

French Treebank For our source treebank, we use the French GSD treebank (Guillaume et al., 2019)¹ as it is sufficiently large in size and predominantly consists of news articles, which aligns better with the newly created MC treebank.

MC Treebank The MC treebank consists of news and blog articles written in Martinican Creole by native speakers. Texts range from 2004 to 2021 and consist of two primary sources: 1) *Kréyolad*² collections which gather all the article contributions of Jude Duranty to the newspaper *Antilla*³ from 2004 to 2018 and 2) the collective blog *Montray Kréyol*⁴ which contains columns from numerous authors, written in French and various (mostly French-based) creoles. Selected text were annotated by the first author. The fully annotated treebank of MC consists of 240 sentences and a total of 4809 tokens.⁵

Annotation of MC Treebank We tokenized the texts using NLTK Tokenizer⁶ and then annotated for POS information using INCePTION (Klie et al., 2018). INCePTION proposes an automatic POS tagger training on the annotations one makes synchronously and retrains itself whenever a new word receives a tag. We then used UD Annotatrix (Tyers

¹Experiments training with all French treebanks were computationally more expensive and yielded poorer results.

²<https://www.potomitan.info/duranty/kreyolad.php>

³<https://antilla-martinique.com/>

⁴<https://www.montraykreyol.org/>

⁵The treebank will be released in the next UD cycle.

⁶<https://www.nltk.org/api/nltk.tokenize.html> We used the default model (English) since we did not expect any differences in punctuation.

Treebank	Train	Dev	Test
FR-GSD	13 072	1 634	1 634
MC	80	80	80

Table 1: Distribution in Train/Dev/Test sets of FR-GSD and Martinican Creole (MC) treebanks.

et al., 2017) for the dependency annotations. The treebank is not annotated for lemmas or morphological information.

5.2 Experimental Setup

Data Splits Due to the small size of the MC corpus, we split the treebank into equal size folds for train, dev, and test of 80 sentences. For more generalized results, we generate three different randomized splits and report results averaged over the three runs. For the French GSD treebank, we use the standard train/dev/test split, unless otherwise noted. Table 1 shows the sizes of the different data sets.

Parser We use the Deep Biaffine parser (Dozat and Manning, 2017) implemented in the SuPar parsing library.⁷ The parser is a neural graph-based dependency parser which uses biaffine attention and biaffine classifier in combination with dimension reducing MLP layers to reduce non-relevant information.

We experiment with different input embeddings: character, POS tag, and BERT embeddings. Note that SuPar always includes word embeddings, so that we can only use (word+)POS and (word+)BERT. For all POS embeddings, we use gold POS tags. For the BERT embeddings, we use the French camemBERT (Martin et al., 2020)⁸.

In addition, we also use a multi-task learning parser where each treebank is treated as a separate task (Sayyed and Dakota, 2021). Both input embeddings into the BiLSTM and the subsequent MLP layers are shared, which allows for information transfer during joint optimization between the treebanks. We also experiment with weighting treebanks with respect to their joint loss contribution, which has shown to be beneficial when data imbalances exist between treebanks (Dakota et al., 2021), as in our case. Results reported are using the scorer from CoNLL2018 shared task (Zeman et al., 2018).

⁷<https://github.com/yzhangcs/parser>

⁸We also experimented with the other large French LM, FlauBERT (Le et al., 2020), but this yielded worse performance

Train	Embed.	UAS	LAS
French	char	25.05	11.73
	POS	65.08	51.95
	BERT	38.23	21.63
MC	char	62.89	48.36
	POS	71.71	62.86
	BERT	63.36	49.83
FR+MC	char	72.95	60.57
	POS	80.75	71.77
	BERT	72.17	58.57

Table 2: Baselines for training on French, Martinican (MC), and concatenated French+Martinican (FR+MC).

6 Results

6.1 Baselines

We first need to establish the baselines, i.e., training on the French training set, training on the Martinican Creole training set, and concatenating these two. Here, we optimize and test on the MC dev set.

Table 2 shows the results for these baseline models. These results show that the French training data gives us the lowest results. The best model, using POS embeddings, results in an LAS of 51.95. Using character and BERT embeddings results in considerable losses (LAS: 11.73 and 21.63); this can be attributed to the significant differences in spelling between French and MC (see section 4). Training on 80 MC sentences is surprisingly successful. Again, using the POS embeddings shows the best results (LAS: 62.86). It is worth noting how beneficial the use of POS embeddings is for MC compared to subword information. One reason is simply the small data size of the MC treebank; another reason may be that some of the linguistic properties of MC are disambiguated via POS tags but not via characters. However, the concatenation of both training sets results in the highest scores overall, with an LAS of 71.77 for POS embeddings. This is particularly interesting given that the French training size is about 136 times the size of the MC training but this small amount is enough to direct the French-trained model in a beneficial direction.

6.2 Optimization

Since we operate in a very low-resource setting, the next question is whether it is worth annotating sentences to use for optimizing the parser or whether the neural architecture does not require target language specific optimization. Thus, we compare a

Dev	Embed.	UAS	LAS
French	char	20.03	9.16
	POS	58.17	45.54
	BERT	33.07	18.38
MC	char	25.05	11.73
	POS	65.08	51.95
	BERT	38.23	21.63

Table 3: MC test performance when optimizing on French and MC.

Dev.	Embed.	Finet.	UAS	LAS
French	char	no	20.03	9.16
		yes	20.03	9.16
	POS	no	58.17	45.54
		yes	58.17	45.54
	BERT	no	33.07	18.38
		yes	33.07	18.38
MC	char	no	25.05	11.73
		yes	64.71	46.83
	POS	no	65.08	51.95
		yes	72.87	60.83
	BERT	no	38.23	21.63
		yes	67.10	49.41

Table 4: Performance with and without fine-tuning on MC.

setting trained and optimized on French with a setting where we train on French and optimize on 80 MC sentences. The results are shown in Table 3.⁹

Our results show increases when the source French model has been optimized on MC as compared to French. This is true for all types of embeddings. The POS model shows a sizable improvement of more than 6 percent points for LAS while the improvements for the character and BERT models are more modest, around 2-3 percent points. However, we do not reach the MC baseline in any setting.

6.3 Fine-tuning

We next experiment with transfer learning in order to see if we can improve on the French baseline by fine-tuning on the MC training set. Given the difference in size, the MC data should not have a noticeable effect, but since the concatenation baseline proved so successful, we need to determine whether fine-tuning on the MC training has the

⁹Note that the results of the French model optimized on MC are repeated from Table 2.

same effect. When training on French, we have two settings: We either optimize on French or on MC. When fine-tuning on MC, we optimize on MC.

Table 4 shows the results of these experiments. Note that the results without fine-tuning are repeated from Table 3. The results show very clearly that fine-tuning is only successful when the first stage is optimized on MC. If we optimize that stage on French, fine-tuning does not result in any improvement. This is likely due to the fact that training a fully optimized French model results in overfitting, which in turns does not allow the little MC data to effectively update the parameters.

When optimizing on MC, we note that all models show a drastic improvement in performance, especially for the BERT and character embeddings. Out of the three types of embeddings, the model using character embeddings benefits the most from fine-tuning, improving from 25.05 to 64.71 for UAS and from 11.73 to 46.83 for LAS, followed by the BERT embeddings model going from 38.23 to 67.10 for UAS and 21.63 to 49.41 for LAS. The most successful model, using POS embeddings, reaches an LAS of 60.83. While this is still below the concatenation model, it shows again the usefulness of POS embeddings.

6.4 Overfitting

One reason for the lack of improvement of the model optimized on French in Table 4 may be that training a fully optimized French model results in overfitting, which in turn does not allow the little MC data to effectively update the parameters. To investigate the issue of overfitting, we perform experiments where we stop the training early. Since it is unclear how to determine good stopping points, we stopped the training at epoch 1 as well as at the 1/4, 1/2, and 3/4 of the optimal number of epochs when using the French model (trained and optimized on French) and perform fine-tuning experiments in two settings, fine-tuning the model on MC, and on the concatenated FR+MC treebank. In both cases, we optimize on MC. The results of these experiments are shown in Table 5.

When comparing between the two fine-tuning settings, we note that none of the 1/4, 1/2, 3/4 or the fully optimized models improve from the MC or MC+FR data during fine-tuning, as results are not substantially different from the ones without fine-tuning. This indicates that the more a model

FT	Emb.	Epoch 1	1/4	1/2	3/4	Full
MC	char	41.59	9.41	9.38	9.32	9.16
MC	POS	36.82	49.81	48.79	47.23	45.54
MC	BERT	11.55	20.30	19.25	18.03	18.38
FR+MC	char	54.17	9.41	9.38	9.32	9.16
FR+MC	POS	65.33	49.81	48.79	47.23	45.54
FR+MC	BERT	11.55	20.30	19.25	18.03	18.38

Table 5: LAS when training on 1/4, 1/2, 3/4 of the best epoch of the French model, fine-tuned on MC or FR+MC.

Embed.	No weight		Weight	
	UAS	LAS	UAS	LAS
char	70.23	56.58	71.44	58.26
POS	64.67	50.46	64.99	50.12
BERT	69.39	56.33	70.76	56.78

Table 6: Results for MTL with non-weighted and weighted losses on the MC task. All weighted experiments use 0.9 for French and 0.1 for MC.

is trained and optimized on French, the less it is able to profit from having access to MC data. The only models showing noticeable benefit from fine-tuning are the epoch 1 character model fine-tuned on MC and the epoch 1 character and POS models fine-tuned on FR+MC, but both are still below their respective baselines.

When we look at the experiments with fewer epochs, we see a deterioration of the results from fewer epochs to the full number of epochs, showing clear signs of overfitting. This trend holds across all conditions but is strongest for the highest performing model using POS embeddings. Here the LAS decreases from 49.81 to 45.54. However, even the results at 1/4 epochs are far below the MC baseline.

6.5 Multi-task Learning

Another approach for information sharing is to use multi-task learning (MTL). By treating each treebank as a task, it allows them to be optimized jointly but does so by combining information with the other treebank in the process rather than sequentially as in a typical transfer learning setup. For this experiment, we have two settings, one without weighting losses, and one with loss weighting. Reducing the weights for the smaller treebank may help reduce the negative transfer that can occur. Given the small size of the MC training set, its contribution to the overall loss may be too high, leading the parser in a sub-optimal direction. This assumption has been shown to hold for a domain adapta-

tion setting (Dakota et al., 2021), where assigning higher weights to the larger and lower weights to the smaller treebank yielded the best performance. Consequently, we assign a loss of 0.9 to the French treebank and 0.1 to the MC treebank.

The results of this experiment are shown in Table 6. Results are generally better than for the fine-tuning setting. However, the best result so far is still the baseline trained on only 80 sentences of MC and using POS (see Table 2), as none of the MTL settings reach this result. When we compare the weighted and non-weighted settings, we see an improvement of about 1.5 points (LAS: from 56.58 to 58.26) for the character model and a minimal gain for the BERT model (LAS: from 56.33 to 56.78), but a small decrease for the POS model. It is noticeable that using POS information leads to substantially worse results in comparison to the other models, thus contradicting the trends of previous experiments. This further re-enforces the notion of negative transfer when sharing POS information.

We next look at a setup where we use the FR+MC concatenated treebank as one of our tasks and the MC treebank as the other, with both optimizing on the same development set, but using different weights.^f

Table 7 shows the results of this experiment (the FR/MC setting is repeated from Table 6). We see that using the combined FR+MC training set gives us a moderate boost of 2-3 percent points over the FR/MC setting. Here, the UAS improves over the best MC-only baseline, the LAS does not. Additionally, we can see that even further reducing the MC weights tends to yield better performance for LAS, suggesting that as the data imbalance becomes extreme, so does the need to downweight the smaller treebank.

Embeddings	Weights		UAS	LAS
	FR	MC		
char	0.90	0.10	71.44	58.26
POS	0.90	0.10	64.99	50.12
BERT	0.90	0.10	70.76	56.78
	FR+MC	MC		
char	0.90	0.10	73.13	59.84
POS	0.90	0.10	69.56	55.81
BERT	0.90	0.10	73.37	60.08
char	0.95	0.05	74.08	61.26
char	1.0	0.01	73.54	61.47
POS	0.95	0.05	70.77	57.66
POS	1.0	0.01	70.71	57.53
BERT	0.95	0.05	72.97	59.82
BERT	1.0	0.01	73.15	60.04

Table 7: Results for the MC task using varying weights for the MTL parser, training on either FR and MC or on FR+MC and MC and testing on MC.

7 Analysis

All the experiments described above tell us that the best method to parse Martinican Creole given a very small treebank is to concatenate the two treebanks. It is unclear why first training on French and then fine-tuning on MC does not result in a similar performance. And it is equally unclear why the POS embeddings are successful in transfer learning but not in a multi-task learning setting. We assume that the two are related and will analyze the data to shed light on these questions.

7.1 Correlation of POS Tags and Parser Errors

We first look into the correlation between specific part of speech tags and parser errors. More specifically we look at label accuracy of incoming arcs per POS tag, with a primary focus on the experiments that include the concatenated FR+MC data during training.

Table 8 presents the results¹⁰ for the FR+MC embeddings baselines and their best respective MTL settings from Table 7. We see the same trends across most open and closed class POS tags within one setting. Since the lexical models (char and BERT) show significantly lower results than the POS model, this points to a disconnect on the lexical level (caused by the different spelling systems) that can only be overcome by adding POS infor-

¹⁰All numbers are averaged over the three folds.

mation. However, in this case, we would expect a better performance of the POS model in our MTL task in comparison to the MC baseline. Since this does not happen, we assume that there are significant differences on the POS level between the two languages, causing negative transfer for the MTL model (one facet of this will be investigated in more detail in the next section.)

One notable trend is related to the accuracies for adjectives and adverbs: While the baseline POS model can parse those POS very successfully, the MTL POS model reaches accuracies that are below the MTL character and BERT models for adjectives and comparable for adverbs (again, see below for an explanation).

7.2 POS Distribution

We now have a closer look at the POS distributions between French and MC, to determine whether these ambiguity rates can give us insights into the differences between French and MC on the POS level. However, a direct comparison does not seem to be feasible since the MC treebank is too small to give us a stable picture, especially compared to the large French treebank. For this reason, we decided to use the full 240 sentences of the MC treebank and to randomly sample 240 sentences from the French treebank (averaged over 10 repetitions). While the small number will introduce some variability, the results will be more comparable across the languages.

When looking at the percentage of ambiguous words, 2.2% of the word types (in the POS lexicon) for French and 7.0% for MC are ambiguous, showing that about 3 times more MC words are ambiguous. Additionally, the percentage of ambiguous word types amounts to 13.0% when we concatenate the French and MC treebanks.

Table 9 shows the rates of ambiguous word types per POS tag. A comparison of French and MC shows that for all POS tags, the MC words are ambiguous about 3 times more often. And while French subordinating conjunctions (SConJ) and prepositions (Adp) tend to be frequently ambiguous, this ratio increases to more than 50% for MC. Additionally, the percentages for the combined treebank shows that the ambiguities are mostly additive, i.e., there is not much overlap between the ambiguous words in French and MC. This at least partly explains the difficulties of the POS models. The most extreme cases are subordinating conjunctions

	Noun	Verb	Adj	Propn	Adv	CConj	SConj	Adp	LAS
baseline char	56.76	63.67	50.36	60.82	57.38	66.80	60.27	68.49	60.57
baseline POS	70.52	72.03	84.45	66.78	87.21	88.12	91.34	89.23	71.77
baseline BERT	54.92	58.47	53.05	58.15	51.66	62.57	55.92	65.44	58.57
MTL char	58.03	64.45	50.90	58.53	62.64	71.08	65.38	66.55	61.47
MTL POS	53.00	58.61	45.35	56.93	59.72	68.49	65.01	68.56	57.66
MTL BERT	55.67	60.96	55.68	60.82	58.45	69.48	58.19	65.20	60.08

Table 8: Accuracy of dependency labels per POS tag for FR+MC baseline and best MTL experiments.

	Noun	Verb	Adj	Propn	Adv	CConj	SConj	Adp	Total
French	2.2%	3.5%	5%	0.5%	8.9%	8.4%	48.8%	18.6%	2.2%
MC	5.7%	9.5%	17.9%	1.7%	24.5%	37.5%	61.8%	51.8%	7.0%
FR+MC	14.5%	15.3%	34.6%	17.3%	35.8%	51.7%	70.6%	62.9%	13.0%

Table 9: Average of ambiguous word types per POS tag.

and prepositions, for which more than 50% are ambiguous in MC and more than 60% in the combined treebank. For the open class POS tags, adjectives and adverbs are the most affected. In the case of adjectives, the combined treebank shows a doubling of the ambiguity rate from MC to the combined treebank, thus indicating not only an increase in ambiguity in MC, but also a high number of words that are only considered adjectives in one of the languages but not both. This partly explains the low results for adjectives in the MTL POS setting in Table 8.

7.3 Example

Martinican creole shows a systematic ambiguity between nouns and adjectives. The word *politik* is one example, as shown in examples (6) and (7). In example (6) the word is misidentified as a noun, which leads the character model to interpret it as an *nmod* of *désizion* instead of its *amod* (see Figure 1). Having access to the gold POS tags in the POS model helps this model disambiguate it correctly.

- (6) zot wè ni an désizion politik
 2PL see there-is a decision political
 “You saw that there is a political decision.”
- (7) fanm an politik
 women in politics

8 Conclusion and Future Work

In this study, we built a first parser of Martinican Creole using French as the supporting language to address the extremely low-resource setting of the creole.

Our main finding is that, surprisingly, we obtain

the best parsing results with our baseline model trained on a concatenation of the French and MC training sets. The success of the concatenated baseline model shows that even with as little as 80 MC sentences in the training set, the POS model is able to direct itself in the right direction.

Even the baseline POS model trained on 80 MC sentences outperforms all transfer and MTL models, the single exception being the UAS of the MTL character and BERT models. Partial explanations for these results can be found in the different spelling systems used for French and MC (see Section 4) and in the high level of ambiguity of MC, and specifically MC adjectives and adverbs. Whether POS tags are needed in neural dependency parsing is still an open question (Anderson and Gómez-Rodríguez, 2020; Zhou et al., 2020), and our findings further complicate this picture. In our case, they can reduce ambiguity in our baselines, but increase ambiguity across the two languages. Since we use gold POS tags, there remains the open question whether the same effects will occur with automatically annotated POS tags.

Our results partially contradict findings by Wang et al. (2019) for Singlish. They also found that in the low resource setting (using 900 Singlish sentences), treebank concatenation outperforms MTL. However in their work, MTL outperformed the baselines for both individual treebanks while we did not see this increase in performance across experiments. Our findings thus confirm that to improve our performances on parsing MC using French, we will need to reduce the imbalance between the two languages, by augmenting the MC data. For the future, we are planning to investi-

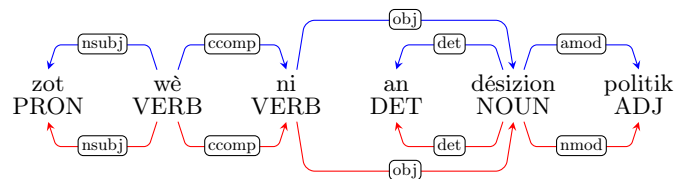


Figure 1: *zot wè ni an désizion politik* parses for POS and char predictions.

gate whether a larger MC training set will have a positive effect in the MTL setup.

However, the fact that as little as 240 annotated sentences, provided that we concatenate them with French data, can yield an LAS in the low 70es indicates that it is possible to develop parsing models for French-based creoles without extensive annotation projects.

References

- Željko Agić, Anders Johannsen, Barbara Plank, Héctor Martínez Alonso, Natalie Schluter, and Anders Søgaard. 2016. Multilingual Projection for Parsing Truly Low-Resource Languages. *Transactions of the Association for Computational Linguistics*, 4:301–312.
- Mark Anderson and Carlos Gómez-Rodríguez. 2020. On the frailty of universal POS tags for neural UD parsers. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 69–96, Online.
- Jean Bernabé. 2004. Éléments d’écologie linguistique appliqués à la situation martiniquaise. In *Créoles-Langages et Politiques linguistiques: Actes du XXVI e Colloque International de Linguistique Fonctionnelle-30 septembre-7 octobre 2002 à Gosier (Guadeloupe)*, pages 13–29. Peter Lang.
- Irshad Bhat, Riyaz A. Bhat, Manish Shrivastava, and Dipti Sharma. 2018. Universal Dependency parsing for Hindi-English code-switching. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 987–998, New Orleans, LA.
- Anders Björkelund, Agnieszka Falenska, Xiang Yu, and Jonas Kuhn. 2017. IMS at the CoNLL 2017 UD shared task: CRFs and perceptrons meet neural networks. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 40–51, Vancouver, Canada.
- Bernard Caron, Marine Courtin, Kim Gerdes, and Sylvain Kahane. 2019. A Surface-Syntactic UD Treebank for Naija. In *TLT 2019, Treebanks and Linguistic Theories, Syntaxfest*, Paris, France.
- Wanxiang Che, Jiang Guo, Yuxuan Wang, Bo Zheng, Huaipeng Zhao, Yang Liu, Dechuan Teng, and Ting Liu. 2017. The HIT-SCIR system for end-to-end parsing of Universal Dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 52–62, Vancouver, Canada.
- Daniel Dakota, Zeeshan Ali Sayyed, and Sandra Kübler. 2021. Bidirectional domain adaptation using weighted multi-task learning. In *Proceedings of the 17th International Conference on Parsing Technologies and the IWPT 2021 Shared Task on Parsing into Enhanced Universal Dependencies (IWPT 2021)*, pages 93–105, Online.
- Ayan Das, Affan Zaffar, and Sudeshna Sarkar. 2017. Delexicalized transfer parsing for low-resource languages using transformed and combined treebanks. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 182–190, Vancouver, Canada.
- Timothy Dozat and Christopher Manning. 2017. Deep biaffine attention for neural dependency parsing. In *5th International Conference on Learning Representations (ICLR 2017)*, Toulon, France.
- GEREC. 1982. *Charte culturelle créole: se pwan douvan avan douvan pwan nou!* GERIC, Centre universitaire Antilles-Guyane.
- Bruno Guillaume, Marie-Catherine de Marneffe, and Guy Perrier. 2019. Conversion et améliorations de corpus du français annotés en Universal Dependencies. *Revue TAL*, 60(2):71–95.
- Marie-Christine Hazaël-Massieux. 2002. Les créoles à base française: une introduction. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d’Aix-en-Provence (TIPA)*, 21:63–86.
- Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9, Santa Fe, New Mexico.
- Hang Le, Loïc Vial, Jibril Frej, Vincent Segonne, Maximin Coavoux, Benjamin Lecouteux, Alexandre Al-lauzen, Benoît Crabbé, Laurent Besacier, and Didier Schwab. 2020. FlauBERT: Unsupervised language

- model pre-training for French. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 2479–2490, Marseille, France.
- Heather Lent, Emanuele Bugliarello, Miryam de Lhoneux, Chen Qiu, and Anders Søgaard. 2021. On language models for creoles. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 58–71, Online.
- William Lewis. 2010. Haitian Creole: How to build and ship an MT engine from scratch in 4 days, 17 hours, & 30 minutes. In *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*, Saint Raphaël, France.
- Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamel Seddah, and Benoît Sagot. 2020. CamemBERT: a tasty French language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online.
- Ailsa Meechan-Maddon and Joakim Nivre. 2019. How to parse low-resource languages: Cross-lingual parsing, target language annotation, or both? In *Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)*, pages 112–120, Paris, France. Association for Computational Linguistics.
- Alice Millour and Karën Fort. 2018. Krik: First Steps into Crowdsourcing POS tags for Kréyòl Gwadeloupéyen. In *CCURL 2018*, Miyazaki, Japan.
- Zeeshan Ali Sayyed and Daniel Dakota. 2021. Annotations matter: Leveraging multi-task learning to parse UD and SUD. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3467–3481, Online.
- Tianze Shi, Felix G. Wu, Xilun Chen, and Yao Cheng. 2017. Combining global models for parsing Universal Dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 31–39, Vancouver, Canada.
- Francis M. Tyers, Mariya Sheyanova, and Jonathan North Washington. 2017. UD Annotatrix: An annotation tool for Universal Dependencies. In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*, pages 10–17, Prague, Czech Republic.
- Clara Vania, Yova Kementchedjheva, Anders Søgaard, and Adam Lopez. 2019. A systematic comparison of methods for low-resource dependency parsing on genuinely low-resource languages. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1105–1116, Hong Kong, China. Association for Computational Linguistics.
- Erik Velldal, Lilja Øvrelid, and Petter Hohle. 2017. Joint UD parsing of Norwegian Bokmål and Nynorsk. In *Proceedings of the 21st Nordic Conference on Computational Linguistics*, pages 1–10, Gothenburg, Sweden.
- Georges Daniel Véronique. 2020. Les langues créoles dans la république française: entre démarcation et revendication1. *Entre francisation et démarcation.: Usages hérités et usages renaissantistes des langues régionales de France*, page 201.
- Hongmin Wang, Jie Yang, and Yue Zhang. 2019. From genesis to creole language: Transfer learning for Singlish Universal Dependencies parsing and POS tagging. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 19(1).
- Hongmin Wang, Yue Zhang, GuangYong Leonard Chan, Jie Yang, and Hai Leong Chieu. 2017. Universal Dependencies parsing for colloquial Singaporean English. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 1732–1744, Vancouver, Canada.
- Daniel Zeman, Jan Hajič, Martin Popel, Martin Potthast, Milan Straka, Filip Ginter, Joakim Nivre, and Slav Petrov. 2018. CoNLL 2018 shared task: Multilingual parsing from raw text to Universal Dependencies. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–21, Brussels, Belgium.
- Houquan Zhou, Yu Zhang, Zhenghua Li, and Min Zhang. 2020. Is pos tagging necessary or even helpful for neural dependency parsing? In *Natural Language Processing and Chinese Computing*, pages 179–191, Cham. Springer International Publishing.