Cross-Lingual Transfer and Structural Complementarity in Low-Resource Creole Dependency Parsing

Anonymous ACL submission

Abstract

We investigate transfer strategies for dependency parsing in two French-based Caribbean Creoles: Haitian Creole (HC) and Martinican Creole (MC). We compare multilingual augmentation, multi-task learning (MTL), pretrained encoder representations, and recurrent versus transformer-based architectures under extremely low-resource conditions. Our results show that transfer effects are highly non-uniform across languages and syntactic phenomena. French augmentation yields the strongest overall improvements, particularly for HC, while sibling-Creole transfer remains competitive despite dramatically smaller auxiliary resources. We further show that concatenation-based augmentation and MTL specialize differently across the syntactic error space: BiLSTM-based multilingual augmentation primarily improves locally constrained dependencies, whereas transformer-based MTL contributes more strongly to coordination and long-distance structures. Fine-grained oracle analyses reveal substantial complementarity across transfer regimes and encoder architectures, indicating that different models capture partially distinct structural generalizations rather than converging toward identical solutions. These findings motivate future fusion pipelines and adaptive routing strategies capable of dynamically leveraging complementary syntactic strengths across models.

1 Introduction

Recent advances in multilingual dependency parsing have demonstrated that syntactic transfer across languages can substantially improve performance in low-resource settings, particularly when leveraging pretrained multilingual encoders and cross-lingual training strategies (Ammar et al., 2016; Kondratyuk, 2019). However, most transfer-based parsing research has focused on languages represented by relatively large treebanks or well-supported multilingual pretraining corpora, leaving extremely low-resource languages underexplored.

French-based Caribbean Creoles such as Haitian Creole (HC) and Martinican Creole (MC) present a particularly challenging setting for multilingual transfer. Although both languages share substantial lexical overlap with French, they also exhibit important syntactic and morphosyntactic divergences shaped by distinct contact ecologies and language histories. As a result, lexical proximity to a high-resource lexifier language does not necessarily guarantee effective syntactic transfer, particularly under extremely low-resource conditions. At the same time, HC and MC share many structural similarities with one another despite their independent sociolinguistic trajectories, raising the question of whether sibling-Creole transfer can compensate for dramatically smaller auxiliary resources.

Despite recent growth in multilingual NLP, syntactic resources for Creole languages remain extremely limited. While HC has recently received initial Universal Dependencies (UD) treebanks and MC has begun to receive UD-style annotation efforts, both languages remain severely under-resourced relative to most multilingual parsing benchmarks. This creates a useful comparative setting for studying how typological proximity, auxiliary data quantity, encoder architecture, and transfer mechanism interact in low-resource parsing.

In this paper, we systematically investigate transfer strategies for dependency parsing in HC and MC. We compare multilingual augmentation, multi-task learning (MTL), pretrained encoder representations, and lexical initialization across recurrent and transformer-based architectures. We further examine how different transfer mechanisms affect distinct regions of the syntactic error space through fine-grained analyses of dependency relations, dependency distance, and model complementarity.

We address the following research questions:

1. To what extent can dependency parsing for extremely low-resource Creole languages be improved through cross-lingual transfer from either a high-resource lexifier language (French) or a sibling Creole language?

2. How do data quantity and typological proximity trade off in low-resource transfer settings? In particular, can closely related low-resource languages provide competitive syntactic transfer despite dramatically smaller auxiliary resources?

3. How do different transfer mechanisms—multilingual augmentation, multi-task learning, and pretrained encoder representations—affect distinct regions of the syntactic error space?

4. Do different transfer regimes and encoder architectures learn complementary syntactic generalizations, and can these complementary strengths be exploited through structured model fusion?

2 Related Work

2.1 Neural Dependency Parsing

Graph-based neural dependency parsers based on biaffine attention have become a standard architecture for syntactic parsing in the Universal Dependencies (UD) framework (Dozat and Manning, 2016). These models combine contextual encoders with task-specific multilayer perceptrons to jointly predict head–dependent arcs and dependency relations. Their modular design has enabled the incorporation of both pretrained contextual representations and auxiliary linguistic features such as part-of-speech tags and character-level embeddings, which have been shown to improve robustness in morphologically rich or low-resource languages (Ballesteros et al., 2015; Straka and Straková, 2017). The integration of large pretrained transformer encoders into biaffine parsing architectures has further shifted the field toward multilingual and low-resource transfer settings, where contextualized representations can compensate for limited annotated syntactic data (Kondratyuk, 2019).

2.2 Domain-Adaptive Pretraining and Fine-Tuning

A major challenge in low-resource dependency parsing concerns the limited availability of annotated treebanks and the difficulty of learning robust syntactic representations from extremely small datasets. Recent work has therefore explored several complementary strategies for improving syntactic transfer under low-resource conditions. One approach relies on pretrained multilingual language models such as mBERT (Devlin et al., 2019) and XLM-RoBERTa (Conneau et al., 2020). While these models have substantially improved multilingual parsing performance (Kondratyuk and Straka, 2019), they often exhibit degraded performance on typologically underrepresented or low-resource languages due to vocabulary mismatches and limited pretraining exposure (Pires et al., 2019). Continued in-domain pretraining or domain-adaptive fine-tuning has therefore emerged as an effective strategy for improving downstream syntactic performance in underrepresented languages (Gururangan et al., 2020). This is especially relevant for Creole languages, where Creole-adapted multilingual encoders such as CreoleVal (Lent et al., 2024) may better capture orthographic, lexical, and morphosyntactic regularities absent from general-purpose multilingual models.

2.3 Cross-Lingual Transfer and Multilingual Augmentation

Another major strategy for low-resource dependency parsing focuses on cross-lingual transfer and multilingual augmentation. Proposed techniques include annotation projection (Hwa et al., 2005), delexicalized transfer (Zeman and Resnik, 2008), multilingual parsing architectures (Ammar et al., 2016), and pooled multilingual training using typologically or genealogically related treebanks (Dabre et al., 2017; Schlichtkrull and Søgaard, 2017; Ahmad et al., 2019; Snæbjarnarson et al., 2023). These approaches expose parsers to a broader range of syntactic structures and lexical distributions while avoiding some of the alignment constraints required by projection-based methods.

2.4 Multi-Task Learning

A third strategy involves multi-task learning (MTL), where parsers share encoder representations across related tasks or languages while maintaining task-specific decoding components (Caruana, 1997; Sayyed and Dakota, 2021; Zhang et al., 2023). In multilingual dependency parsing, hard parameter sharing has proven effective in several low-resource settings, although shared encoders may also introduce substantial negative transfer

when syntactic conventions or data distributions diverge across languages (Ruder, 2019).

2.5 Creole Dependency Parsing

Recent work on Creole dependency parsing has shown that multilingual augmentation from related languages can substantially improve parsing performance even under extremely small treebank conditions, with simple treebank concatenation often outperforming more complex transfer and multi-task configurations (Mompelat et al., 2022). These earlier experiments further revealed substantial instability across encoder types and transfer regimes, particularly under multi-task learning and severe data imbalance conditions.

The present study builds directly on these findings by expanding both the experimental scope and the granularity of analysis. In addition to evaluating newer pretrained multilingual encoders and broader transfer configurations across Haitian and Martinican Creole, we systematically investigate how different transfer mechanisms affect specific regions of the syntactic error space. Rather than relying exclusively on aggregate parsing metrics, we examine transfer behavior across dependency relations, dependency distance, encoder architectures, and model complementarity in order to better understand the structural dynamics underlying low-resource multilingual transfer.

3 Data and Methodology

3.1 Corpora

We evaluate dependency parsing performance on two French-based Caribbean Creole languages: Haitian Creole (HC) and Martinican Creole (MC). HC has recently seen the development of Universal Dependencies (UD) resources annotated according to the UD framework (De Marneffe et al., 2021; Jagodzińska et al., 2024), while MC remains largely absent from existing syntactic NLP infrastructures. Despite their lexical relationship to French, HC and MC remain extremely low-resource in terms of syntactic annotation. The HC-Autogramm treebank used in this study contains approximately 144 manually annotated sentences (Jagodzińska et al., 2024), while the MC treebank consists of 236 manually annotated sentences collected and annotated as part of our ongoing work on MC dependency parsing.

Recent efforts have further expanded UD coverage for HC through the annotation of Biblical texts and related corpora (HC-Adolphe[1]). Although this substantially increases the amount of available annotated data, preliminary inspection revealed a number of annotation inconsistencies and structural analyses that diverge from our annotation decisions and from standard UD practices in ways likely to affect parser training. In addition, the HC-Adolphe corpus consists almost exclusively of automatically annotated Biblical material, producing both domain and stylistic distributions that differ substantially from the more heterogeneous written data represented in HC-Autogramm. Because the HC-Autogramm treebank contains literary, journalistic, and religious texts closer in genre diversity to the MC corpus used in this study, we restrict the present experiments to HC-Autogramm in order to preserve greater annotation and domain comparability across languages.

The MC corpus is composed primarily of blog posts, opinion texts, and news-oriented writings produced by Martinican authors, including Raphaël Confiant and Jid Duranty, originally published on the now-defunct Montray Kréyol platform. The corpus therefore reflects naturally occurring written MC usage across relatively heterogeneous registers.

In addition to HC and MC, we make use of the French GSD UD treebank (Guillaume et al., 2019) as an auxiliary resource for cross-lingual transfer experiments. French serves as a high-resource lexifier language with substantial lexical overlap and partial structural similarity to both Creoles while also differing in important morphosyntactic respects. Following prior work (Mompelat et al., 2022), we select the GSD treebank in part because its size remains large enough to provide meaningful multilingual augmentation without overwhelmingly dominating the substantially smaller HC and MC treebanks during training. The corpus additionally contains journalistic and web-oriented written material that is comparatively closer in domain to the HC and MC corpora than more specialized or literary French UD treebanks.

3.2 Parsing Architecture

All experiments are conducted using the SuPar[2] parsing framework and implement graph-based neural dependency parsers following the biaffine architecture of Dozat and Manning (2016). We

[1] https://github.com/UniversalDependencies/UD_Haitian_Creole-Adolphe

[2] https://github.com/yzhangcs/parser

Language	Train	Dev	Test
HC Autogramm	76	32	36
MC	124	53	59
French GSD	14,449	–	–

Table 1: Dataset splits used in the experiments.

compare two encoder paradigms:

- **Transformer encoder**: contextualized token representations obtained through fine-tuning pretrained transformer language models.
- **BiLSTM encoder**: recurrent contextual representations built from learned word embeddings optionally augmented with POS tags, character-level embeddings, and frozen contextual transformer features.

Multi-task experiments are implemented using a hard parameter sharing architecture in which a single encoder is jointly optimized across multiple language-specific parsing tasks (Sayyed and Dakota, 2021)[3]. Each task retains independent multilayer perceptron and biaffine decoding layers while sharing encoder parameters.

3.3 Evaluation and Experimental Setup

We adopt a staged experimental protocol designed to isolate the effects of different transfer mechanisms. Performance is measured using labeled and unlabeled attachment scores (LAS/UAS) computed with the official CoNLL evaluation script. In addition to standard parsing metrics, we introduce Content-Sensitive LAS (C-LAS), which excludes function-word categories (DET, ADP, AUX, PART, SCONJ, CCONJ) in order to focus evaluation on content-bearing syntactic structure where transfer effects are more variable and structurally informative.

Given the extremely limited size of the HC and MC treebanks, experimental variance across random initialization was substantial. To mitigate instability and reduce dependence on a single split configuration, all experiments were repeated across three random seeds unless otherwise specified, and reported scores correspond to mean performance across runs.

Because HC and MC contain only a few hundred annotated sentences, standard train/dev/test partitioning can disproportionately affect performance estimates. We therefore adopt fixed manually curated splits, as shown in Table 1, designed to preserve a reasonable distribution of dependency structures across partitions while avoiding overlap between near-duplicate or closely related sentences.

3.4 Transfer Configuration

For multilingual augmentation experiments, the French auxiliary treebank consists of approximately 14.4k training sentences. French data was used exclusively as an auxiliary transfer resource and was never included in HC or MC evaluation sets.

Following prior work on multilingual parsing, we consider augmented training configurations formed by concatenating treebanks from related languages. We denote these configurations using shorthand notation of the form $X{+}Y$, where the training data of language X is appended to that of target language Y, while evaluation is always performed on Y. For example, `fr+hc` refers to a model trained on the union of the French and HC treebanks and evaluated on HC.

Experiments proceed in four stages: (1) representation transfer through encoder and feature ablations, (2) data-level transfer through multilingual treebank pooling, (3) encoder-level transfer through multi-task learning, and (4) fine-grained error analysis by dependency relation type and attachment distance. Oracle analyses are used not as achievable performance estimates but as diagnostic probes of structural complementarity across transfer regimes and encoder architectures.

3.5 Encoder and Feature Configurations

Preliminary monolingual experiments were conducted to determine the encoder and feature configurations used in downstream transfer analyses. Across both HC and MC, the CreoleVal encoder substantially outperforms generic multilingual XLM-R and French-specific CamemBERT language models. Transformer-based models using CreoleVal achieve 68.71 LAS for HC and 65.57 LAS for MC, compared to 58.91/58.56 for multilingual XLM-R and 56.12/55.62 for CamemBERT. These results indicate that in-domain pretraining on Creole text is substantially more beneficial than lexifier-language proximity alone. For recurrent models, the combination of POS tags, character embeddings, and frozen contextual representations (`tag+char+bert`) yielded the strongest performance overall, reaching 69.15 LAS for HC and 64.36 for MC. Finally, we evaluate lexical ini-

[3] https://github.com/zeeshansayyed/multiparser

Configuration	UAS	LAS	C-LAS
HC			
HC mono (BiLSTM)	79.96	72.64	67.92
FR+HC (BiLSTM)	**85.34**	**76.80**	69.87
MC+HC (BiLSTM)	82.81	76.76	**71.89**
HC mono (BERT)	76.54	68.34	63.92
FR+HC (BERT)	**79.70**	69.45	64.03
MC+HC (BERT)	79.21	**70.94**	**67.38**
MC			
MC mono (BiLSTM)	74.54	66.46	61.93
FR+MC (BiLSTM)	**77.31**	67.79	64.65
HC+MC (BiLSTM)	76.79	**68.89**	**65.54**
MC mono (BERT)	73.35	64.94	61.36
FR+MC (BERT)	**76.62**	67.42	62.33
HC+MC (BERT)	76.44	**67.53**	**63.96**

Table 2: Cross-lingual augmentation results averaged across three random seeds.

tialization strategies using GloVe and fastText embeddings from HC, French, and English. After controlling for embedding dimensionality, related-language fastText embeddings become competitive, particularly for sibling-Creole transfer. However, differences across lexical initializations remain comparatively small once contextual representations are introduced. GloVe therefore provides the most stable configuration overall and is used for all subsequent BiLSTM experiments.

4 Results

4.1 Data-Level Transfer

We first evaluated whether syntactic knowledge can be transferred through cross-lingual data augmentation by training parsers on pooled treebanks drawn from either French or a sibling Creole language and evaluating performance on the target language. In addition to standard LAS/UAS, we report Content-Sensitive LAS (C-LAS), which excludes tokens whose universal POS tags correspond to primarily functional categories (`DET`, `ADP`, `AUX`, `PART`, `SCONJ`, `CCONJ`) in order to focus evaluation on content-bearing syntactic structure. Because function-word relations are comparatively stable across transfer settings, C-LAS provides a more sensitive measure of structurally informative transfer on content-bearing dependencies.

In Table 2, French augmentation consistently improves parsing performance across both languages and encoder families. For HC, French augmentation increases LAS from 72.64 to 76.80 under the BiLSTM encoder, while MC exhibits gains from 66.46 to 67.79 under the same configuration.

Cross-Creole augmentation also consistently improves performance, indicating that HC and MC share transferable structural regularities despite their limited training data. Under the BiLSTM encoder, the `mc+hc` configuration improves HC LAS from 72.64 to 76.76, while `hc+mc` improves MC LAS from 66.46 to 68.89.

Importantly, Creole-to-Creole augmentation remains remarkably competitive with French-based augmentation despite relying on dramatically smaller auxiliary resources. In HC, sibling-Creole augmentation achieves near-parity with French augmentation in aggregate LAS (76.76 vs. 76.80) while substantially outperforming it on C-LAS (71.89 vs. 69.87). Similar patterns are observed in MC, where HC augmentation slightly exceeds French augmentation under both BiLSTM and BERT encoders. These findings suggest that typological proximity may partially compensate for differences in corpus size and may encode complementary syntactic generalizations not captured by French augmentation alone.

4.2 Encoder-Level Transfer

While concatenation-based augmentation exposes parsers to a broader range of syntactic constructions, it does not explicitly encourage models to learn shared representations across languages. We therefore evaluate multi-task learning (MTL) as a form of encoder-level transfer in which related parsing tasks jointly influence a shared latent representation space through hard parameter sharing.

Multi-task experiments employ a shared encoder jointly trained across HC and MC parsing tasks with task-specific decoding heads. Table 3 summarizes results across monolingual, bilingual, and asymmetric multilingual MTL configurations.

Results reveal strong asymmetries across encoder families. Recurrent MTL configurations degrade sharply under encoder sharing, with direct HC$\times$MC training reducing LAS from 68.86 to 56.50 for HC and from 60.81 to 47.81 for MC. Transformer-based encoders remain substantially more stable under the same configuration, with HC improving slightly (68.53 $\rightarrow$ 69.16 LAS) and MC remaining comparatively robust (68.63 $\rightarrow$ 67.50 LAS). These results suggest that pretrained transformer representations tolerate multilingual encoder sharing considerably better than recurrent architectures trained primarily from scratch.

Several asymmetric configurations further reveal evidence of indirect cross-task transfer. HC improves when paired with an FR-augmented MC task (`hc` $\times$ `fr+mc`), reaching 69.90 LAS despite

Configuration	HC UAS	HC LAS	MC UAS	MC LAS
MTL mono (BiLSTM)	78.55	68.86	72.87	60.81
MTL HC×MC (BiLSTM)	73.53	56.50	66.32	47.81
MTL mono (BERT)	75.98	68.53	75.95	68.63
MTL HC×MC (BERT)	78.06	69.16	75.17	67.50
MTL FR+HC×MC (BERT)	**80.44**	**70.71**	77.40	68.80
MTL HC×FR+MC (BERT)	79.29	69.90	**77.88**	**69.09**

Table 3: Multi-task learning results averaged across three random seeds.

HC itself never directly observing French data within its own task. Similarly, MC achieves strong performance when paired with an FR-augmented HC task (`fr+hc × mc`), reaching 68.80 LAS despite MC remaining monolingual within its own task. These asymmetric gains suggest that syntactic information introduced through one task can propagate indirectly through the shared encoder to benefit another language.

5 Structural Transfer Analysis

To better understand the structural effects of transfer, we perform a fine-grained error analysis across dependency relation types, dependency distance, and encoder architectures. For fine-grained analysis, dependency relations are further grouped into broader functional classes, including core arguments, modifiers, function relations, coordination, and compounds following Universal Dependencies conventions.

We further compute oracle analyses within each experimental pool (MTL vs. non-MTL) in order to measure structured complementarity across models. Oracle analyses are used strictly as diagnostic upper bounds on structural complementarity rather than as deployable ensemble estimates, approximating the maximum achievable performance if the correct prediction could always be selected from a pool of parsers (Hall and Nilsson, 2006; Kay, 2000).

For each dependency category or distance bin, oracle gains correspond to the additional LAS obtainable relative to the strongest individual model if the correct prediction could always be selected from any model within the corresponding pool.

5.1 Relation-Level Effects

Table 4 shows that transfer effects are highly non-uniform across syntactic phenomena. Modifier relations exhibit the largest oracle gains overall, reaching +37.9 points in HC under the non-MTL pool and remaining above +20 points across all experimental settings. In contrast, core argument relations exhibit substantially smaller gains, indicating that most systems converge on similar analyses for canonical predicate–argument structure while diverging more strongly on modifiers, coordination, and structurally ambiguous non-core relations.

Under non-MTL settings, the strongest individual systems are generally BiLSTM-based augmentation models, with `fr+hc_BiLSTM` dominating HC and `hc+mc_BiLSTM` dominating MC. However, oracle gains remain extremely large even within these pools, particularly for modifier relations and mid-range dependencies. This indicates that French-augmented and Creole-augmented parsers frequently succeed on different structural configurations despite achieving similar aggregate LAS.

5.2 Dependency Distance and Transfer Specialization

Dependency distance reveals a distinct transfer profile. Concatenation-based augmentation primarily benefits short-range dependencies and locally constrained attachment decisions, whereas MTL contributes comparatively stronger gains for mid- and long-distance structures. Under MTL, oracle gains exceed +28 points for HC dependencies spanning 3–5 tokens, while MC long-distance dependencies above 10 tokens reach gains above +35 points across both MTL and non-MTL pools.

5.3 Encoder Complementarity

Strong asymmetries also emerge across encoder families. BiLSTM systems dominate most concatenation-based augmentation settings, particularly for locally constrained structures, while transformer-based systems remain substantially more stable under MTL and exhibit stronger complementarity on coordination, modifier attachment, and long-distance dependencies. In particular, asymmetric BERT MTL configurations consistently outperform recurrent MTL systems, suggesting that pretrained transformer encoders more effectively leverage latent shared representations across tasks.

Analysis	Phenomenon	HC Non-MTL	HC MTL	MC Non-MTL	MC MTL
Relation-level	Coordination	+7.7	+15.4	**+29.0**	+25.8
	Core arguments	+12.0	+16.7	+11.0	+14.2
	Modifiers	**+37.9**	**+29.0**	+22.2	+20.2
	Function relations	+5.1	+14.9	+9.7	+11.8
Dependency distance	Distance 1	+12.5	+15.3	+11.7	+10.4
	Distance 2	**+24.1**	+21.4	+17.8	+16.9
	Distance 3–5	+17.9	**+28.2**	+27.2	+24.6
	Distance 11+	+17.1	+19.5	**+37.8**	**+35.1**

Table 4: Oracle gains relative to the strongest individual model within each experimental pool. Values indicate the maximum additional LAS obtainable by selecting the correct prediction from any model in the pool.

These findings suggest that low-resource multilingual transfer may be better understood as a distribution of partially specialized syntactic competencies rather than as convergence toward a single optimal multilingual parser. Different augmentation sources, encoder architectures, and transfer regimes can capture complementary structural generalizations. We therefore evaluate whether these complementary strengths can be exploited through model fusion.

6 Model Complementarity and Fusion

Following prior work on parser complementarity and parser integration in dependency parsing (McDonald and Nivre, 2011; Surdeanu and Manning, 2010), we implement a series of diagnostic heuristic routing strategies designed to probe structural complementarity across parser pools. Rather than combining parser probabilities or performing deployable ensemble inference, these analyses estimate how much complementary syntactic information can be recovered by routing dependency predictions according to coarse structural properties such as dependency relation class and attachment distance. We evaluate relation-aware routing, distance-aware routing, and a combined distance-and-relation strategy. Combined routing consistently produced the strongest heuristic performance across all four experimental tracks and is therefore reported in Figure 1.

Figure 1 compares the strongest individual parsers, combined distance-and-relation routing, and oracle upper bounds across all experimental tracks. Heuristic routing yields consistent improvements relative to the best single systems, with the strongest gains emerging under combined distance-and-relation routing (+2.23 LAS for HC Non-MTL, +2.00 for HC MTL, +2.51 for MC Non-MTL, and +1.21 for MC MTL).

Despite these improvements, substantial gaps remain between heuristic routing and oracle upper bounds across all experimental conditions. This indicates that coarse structural routing captures only part of the complementarity observed across transfer regimes. In many cases, models succeed on different structural configurations even within the same dependency class, limiting the effectiveness of static category-level heuristics.

The remaining oracle gap further suggests that different transfer strategies capture complementary structural generalizations that become visible only through fine-grained syntactic analysis. These findings indicate that complementary syntactic information is distributed across augmentation regimes, encoder architectures, and transfer mechanisms rather than concentrated within a single universally optimal parser.

7 Conclusion

We present a systematic evaluation of transfer mechanisms for dependency parsing in Haitian and Martinican Creole, examining representation-level transfer through pretrained encoders, data-level transfer through multilingual augmentation, and encoder-level transfer through multi-task learning.

Our results show that transfer effects are highly non-uniform across syntactic structure. Different transfer regimes specialize in different regions of the dependency space: concatenation-based augmentation primarily benefits modifier attachment and local structure, while multi-task learning contributes more strongly to coordination and long-distance dependencies. These effects further depend on encoder architecture, with recurrent and transformer-based models exhibiting distinct structural strengths and failure modes.

Importantly, sibling-Creole transfer frequently approaches—and occasionally exceeds—the performance of French-based augmentation despite relying on dramatically smaller auxiliary resources.

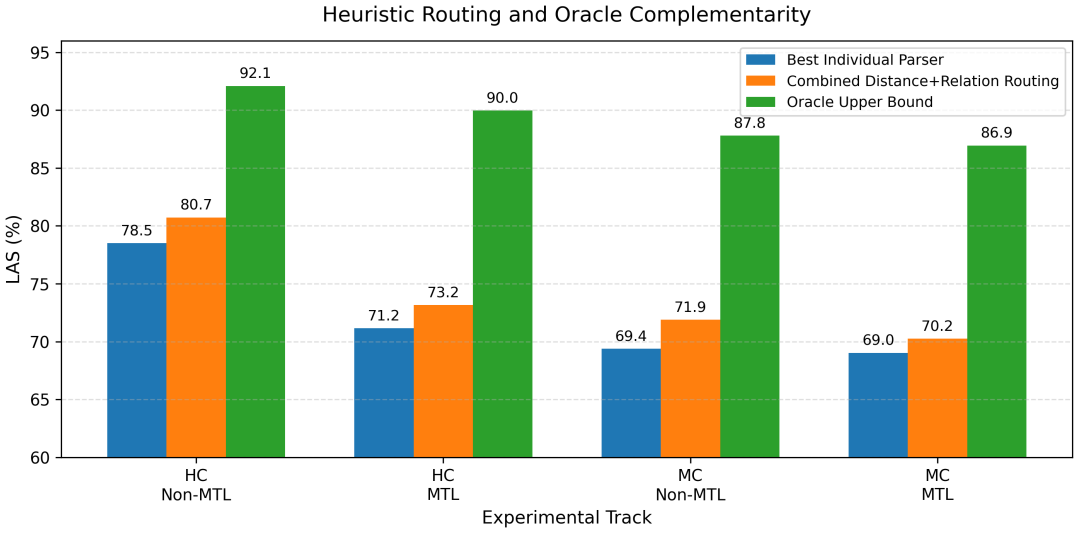

Figure 1: Comparison between the strongest individual parsers, combined distance-and-relation routing, and oracle upper bounds across experimental tracks. While heuristic routing consistently improves over the best single systems, substantial oracle gaps remain across all conditions.

This suggests that typological proximity can compensate for differences in corpus size and that closely related low-resource languages may encode complementary syntactic information not captured by high-resource transfer alone.

Fine-grained error analysis and oracle comparisons further reveal that models trained under different transfer regimes capture distinct syntactic generalizations rather than converging toward identical predictions. This structured complementarity is especially concentrated in modifier relations, coordination, and long-distance dependencies, where oracle gains remain substantial even between high-performing models.

Finally, relation-aware fusion demonstrates that a portion of this complementarity can be operationalized through structurally informed routing strategies, although substantial oracle gaps remain. These findings suggest that low-resource parsing may benefit less from searching for a single universally optimal transfer strategy than from architectures capable of dynamically combining specialized syntactic representations across languages, transfer regimes, and encoder types.

8 Limitations

Several limitations should be considered when interpreting the findings presented in this study.

First, both HC and MC remain extremely low-resource settings with comparatively small treebanks. Although this constraint reflects the real-world conditions faced by many Creole languages, it also increases sensitivity to annotation decisions, train–test splits, and random seed variation. The HC experiments further rely exclusively on the HC-Autogramm treebank, as preliminary inspection of HC-Adolphe revealed annotation inconsistencies and substantial domain differences likely to affect parser training. Consequently, the results presented here may not fully generalize to other Haitian Creole corpora or annotation schemes.

Second, the transfer configurations explored in this work remain limited to French and two French-based Creoles. While this controlled setting allows us to isolate interactions between typological proximity, lexical overlap, and multilingual transfer, the conclusions may not extend directly to unrelated Creole languages or to multilingual settings involving more distant language families.

Third, the fusion and oracle analyses presented in Sections 5 and 6 are diagnostic rather than deployable parsing strategies. Relation-aware routing decisions are derived from empirical performance observed on the evaluation data itself and therefore represent upper-bound analyses of model complementarity rather than realistic inference-time systems. Similarly, oracle scores quantify the maximum achievable performance obtainable from a given parser pool and should not be interpreted as directly attainable parsing accuracy.

Finally, the present study focuses primarily on dependency parsing accuracy and structural error behavior within the Universal Dependencies frame-

work. Although the analyses reveal important asymmetries across transfer mechanisms and syntactic phenomena, they do not directly investigate the internal representations learned by the models or the broader downstream effects of parsing quality on other NLP tasks for Creole languages.

References

Wasi Ahmad, Zhisong Zhang, Xuezhe Ma, Eduard Hovy, Kai-Wei Chang, and Nanyun Peng. 2019. On difficulties of cross-lingual transfer with order differences: A case study on dependency parsing. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2440–2452.

Waleed Ammar, George Mulcaire, Miguel Ballesteros, Chris Dyer, and Noah A Smith. 2016. Many languages, one parser. *Transactions of the Association for Computational Linguistics*, 4:431–444.

Miguel Ballesteros, Chris Dyer, and Noah A Smith. 2015. Improved transition-based parsing by modeling characters instead of words with lstms. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 349–359.

Rich Caruana. 1997. Multitask learning. *Machine learning*, 28(1):41–75.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 8440–8451.

Raj Dabre, Tetsuji Nakagawa, and Hideto Kazawa. 2017. An empirical study of language relatedness for transfer learning in neural machine translation. In *Proceedings of the 31st Pacific Asia conference on language, information and computation*, pages 282–286.

Marie-Catherine De Marneffe, Christopher D Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal dependencies. *Computational linguistics*, 47(2):255–308.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.

Timothy Dozat and Christopher D Manning. 2016. Deep biaffine attention for neural dependency parsing. *arXiv preprint arXiv:1611.01734*.

Bruno Guillaume, Marie-Catherine de Marneffe, and Guy Perrier. 2019. Conversion et améliorations de corpus du français annotés en Universal Dependencies. *Revue TAL : traitement automatique des langues*, 60(2):71–95.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 8342–8360.

Johan Hall and Jens Nilsson. 2006. *CoNLL-X Shared-Task: Multi-lingual Dependency Parsing*. Matematiska och systemtekniska institutionen.

Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Natural language engineering*, 11(3):311–325.

Sandra Jagodzińska, Claudel Pierre-Louis, Sylvain Kahane, Agata Savary, and Emmanuel Schang. 2024. Le premier corpus arboré en créole haïtien. Submitted paper presented at the Journée d'études "Le créole haïtien : histoire, évolution, grammaire et lexique", Université d'État d'Haïti.

Martin Kay. 2000. Guides and oracles for linear-time parsing. In *Proceedings of the Sixth International Workshop on Parsing Technologies*, pages 6–10.

Dan Kondratyuk. 2019. Cross-lingual lemmatization and morphology tagging with two-stage multilingual bert fine-tuning. In *Proceedings of the 16th workshop on computational research in phonetics, phonology, and morphology*, pages 12–18.

Dan Kondratyuk and Milan Straka. 2019. 75 languages, 1 model: Parsing universal dependencies universally. 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 2779–2795.

Heather Lent, Kushal Tatariya, Raj Dabre, Yiyi Chen, Marcell Fekete, Esther Ploeger, Li Zhou, Ruth-Ann Armstrong, Abee Eijansantos, Catriona Malau, and 1 others. 2024. Creoleval: Multilingual multitask benchmarks for creoles. *Transactions of the Association for Computational Linguistics*, 12:950–978.

Ryan McDonald and Joakim Nivre. 2011. Analyzing and integrating dependency parsers. *Computational Linguistics*, 37(1):197–230.

Ludovic Mompelat, Daniel Dakota, and Sandra Kübler. 2022. How to parse a creole: When martinican creole meets french. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4397–4406.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual bert? In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 4996–5001.

Sebastian Ruder. 2019. *Neural transfer learning for natural language processing*. Ph.D. thesis, NUI Galway.

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.

Michael Schlichtkrull and Anders Søgaard. 2017. Cross-lingual dependency parsing with late decoding for truly low-resource languages. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 220–229.

Vésteinn Snæbjarnarson, Annika Simonsen, Goran Glavaš, and Ivan Vulić. 2023. Transfer to a low-resource language via close relatives: The case study on faroese. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 728–737.

Milan Straka and Jana Straková. 2017. Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In *Proceedings of the CoNLL 2017 shared task: Multilingual parsing from raw text to universal dependencies*, pages 88–99.

Mihai Surdeanu and Christopher D Manning. 2010. Ensemble models for dependency parsing: cheap and good? In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 649–652.

Daniel Zeman and Philip Resnik. 2008. Cross-language parser adaptation between related languages. In *Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages*.

Zhihan Zhang, Wenhao Yu, Mengxia Yu, Zhichun Guo, and Meng Jiang. 2023. A survey of multi-task learning in natural language processing: Regarding task relatedness and training methods. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 943–956.