

# Limited-Resource Adapters Are Regularizers, Not Linguists

Marcell Fekete<sup>1</sup> Nathaniel R. Robinson<sup>2</sup> Ernests Lavrinovics<sup>1</sup>  
E. Djeride Jean-Baptiste<sup>3</sup> Raj Dabre<sup>4</sup> Johannes Bjerva<sup>1</sup> Heather Lent<sup>1</sup>

<sup>1</sup>Aalborg University, Denmark; <sup>2</sup>Johns Hopkins University, USA;

<sup>3</sup>University of Ottawa, Canada; <sup>4</sup>IIT Madras, India  
{mrfe, hc1e}@cs.aau.dk

## Abstract

Cross-lingual transfer from related high-resource languages is a well-established strategy to enhance low-resource language technologies. Prior work has shown that adapters show promise for, e.g., improving low-resource machine translation (MT). In this work, we investigate an adapter souping method combined with cross-attention fine-tuning of a pre-trained MT model to leverage language transfer for three low-resource Creole languages, which exhibit relatedness to different language groups across distinct linguistic dimensions. Our approach improves performance substantially over baselines. However, we find that linguistic relatedness—or even a lack thereof—does not covary meaningfully with adapter performance. Surprisingly, our cross-attention fine-tuning approach appears equally effective with randomly initialized adapters, implying that the benefit of adapters in this setting lies in parameter regularization, and not in meaningful information transfer. We provide analysis supporting this regularization hypothesis. Our findings underscore the reality that neural language processing involves many success factors, and that not all neural methods leverage linguistic knowledge in intuitive ways.

## 1 Introduction and Background

For a majority of the world’s languages, data scarcity remains a barrier for reliable machine translation (MT). Cross-lingual transfer from high-resource to low-resource languages has been shown to help mitigate this (Zoph et al., 2016; Tars et al., 2021), but has still not closed the gap between high- and low-resource scenarios. In this work, we seek to exploit methods from parameter-efficient fine-tuning (PEFT) to further improve cross-lingual transfer for a Creoles, an under-served group of low-resource languages.

PEFT methods such as bottleneck adapters (Houlsby et al., 2019; Pfeiffer et al., 2020), LoRA

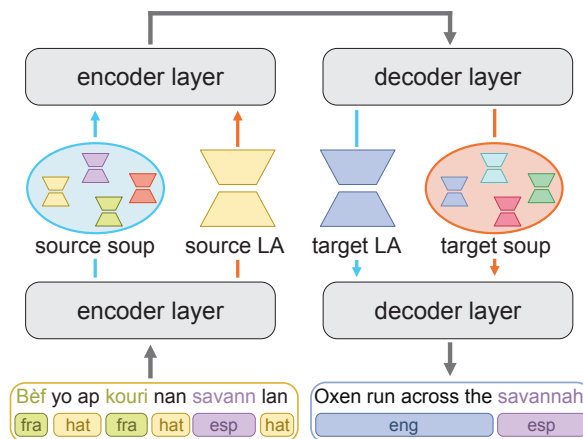


Figure 1: Overview of the MT transfer experiments between English and Creoles. The ↑ path shows the use of source language adapter in the encoder and the ‘souped’ (weight-averaged) target and transfer language adapters in the decoder layers. The ↓ path shows the opposite direction.

(Hu et al., 2021), and (IA)<sup>3</sup> (Liu et al., 2022a) are widely used to adapt language models to low-resource languages (Parović et al., 2022; Parovic et al., 2023; Rathore et al., 2023; Vázquez et al., 2024; Dementieva et al., 2025). This is in part because they are robust to hyperparameter choices and training data amounts, in contrast to full model fine-tuning (Pfeiffer et al., 2024). Some PEFT methods can successfully use monolingual data to improve MT performance; for example, Üstün et al. (2021) train denoising adapters for both source and target languages, inserting the adapters between the encoder and decoder layers of an MT model, respectively. By freezing all model weights except for the cross-attention of the decoder, they then carry out cross-attention fine-tuning (CAFT) with only a limited amount of parallel data. Chronopoulou et al. (2023a) also address data scarcity in adaptation to new domains following Wortsman et al. (2022) by “souping” (i.e., weight-space averaging) pretrained domain adapters, as

in Equation 1 (where  $\theta_i$  represent the weights of each domain adapter). We combine the principles of these prior works in our paper.

$$\theta_{soup} = \frac{1}{l} \sum_{i=1}^l \theta_i \quad (1)$$

Prior work has also catalyzed transfer by training language family adapters to leverage phylogenetic relations between languages (Faisal and Anastasopoulos, 2022; Chronopoulou et al., 2023b). Other factors of selecting transfer languages have been investigated in the past besides phylogeny. These include similarity in terms of explicit grammatical features like word order (Pires et al., 2019), similarity in terms of language representation space (Üstün et al., 2022), as well as lexical overlap (Pfeiffer et al., 2021; de Vries et al., 2022). Other metrics have been proposed to select candidate transfer languages, including Subword Evenness (SuE; Pelloni et al., 2022) which quantifies how uneven subword lengths are in a multilingual tokenizer: uneven implies successful transfer.

In this work we focus on MT for Creole languages, a set of low-resource languages with hundreds of millions of speakers who presently lack reliable MT systems, despite demonstrable need for them (Neto et al., 2020; Robinson et al., 2024). Creole language technologies have not been thoroughly researched or developed; yet several works that do exist inspired our decision to explore them (cf. our literature review in Appendix A). Although most Creoles are genealogically related to high-resource languages (*e.g.*, English, French, Spanish), previous works have had mixed success leveraging transfer learning to boost performance for Creoles (Lent et al., 2022a, 2024; Robinson et al., 2022, 2023). No prior work has conclusively determined which languages to select for successful Creole transfer, explored PEFT in this application, nor compared linguistic and non-linguistic language relations in Creole transfer language selection. These aspects combined make Creoles a viable testbed for comparing transfer criteria.

We apply a new adapter method inspired by Üstün et al.’s (2021) approach to MT for three Creole languages. (See details in §2.) We experimented with this approach to compare a large variety of dimensions of linguistic relatedness and found that, contrary to our initial hypotheses, gains are virtually indistinguishable regardless of which transfer criterion we apply. Still more surprisingly,

incorporating unrelated languages or noise to the Creole adapters yields comparable improvements. We conjecture that even untrained adapters may assist cross-attention fine-tuning by acting as regularizers. We couple this hypothesis with analysis presented in §3.

We contribute both ① a novel regularization method combining adapter souping with cross-attention fine-tuning that surpasses baselines by 8 BLEU in some settings, and ② evidence that some of the benefit of adapters lies in regularization, independent of a wide range of linguistic information.

## 2 Methodology and Experiments

We follow Üstün et al. (2021) using their two-step methodology. First, we train monolingual denoising adapters separately for each language. At inference time, we add the relevant adapters to the encoder and decoder corresponding to the source and target languages, and fine-tune the decoder cross-attention on small amounts of parallel bitexts. Furthermore, we induce transfer by weight-averaging ("souping") Creole adapters with candidate transfer language adapters inspired by previous efforts (Faisal and Anastasopoulos, 2024; Wortsman et al., 2022; Chronopoulou et al., 2023a). The novelty lies in combining source and target language adapters, cross-attention fine-tuning and souping in a single method for improving low-resource MT. (See more details about our methodology in Section 2.3).

We focus our efforts on three Creole languages: Haitian, Papiamentu, and Sango, spoken in total by roughly 18 million speakers in Haiti, the ABC Islands, and the Central African Republic, respectively.<sup>1</sup> This selection is primarily due to the intersection of data availability and support by NLLB-200, the MT model we apply in experiments (NLLB Team et al., 2022). Despite their unique contexts, these languages share commonalities in socio-historical development: they evolved from linguistic contact between European and African languages resulting from Western colonialism (Kouwenberg and Singler, 2009). As such, they can trace parts of their lexicon and grammar back to both European and African languages.

### 2.1 Selecting Languages for Transfer

We examine five principled approaches to select transfer languages, which range from linguistically motivated to purely data-driven. First, we

<sup>1</sup>According to [Ethnologue](#).

Dataset	Domain	Source
MADLAD	Web scrape	Kudugunta et al., 2023
NLLB-OPUS	Web scrape	NLLB Team et al., 2022
FLORES-200	Wikipedia	NLLB Team et al., 2022

Table 1: Datasets and domains used. We use MADLAD to train adapters, NLLB-OPUS to fine-tune cross-attention, and FLORES-200 to evaluate.

leverage language phylogeny, selecting likely Indo-European (**IE**; Rickford and McWhorter, 1997; Lefebvre, 2011) and Niger-Congo (**NC**; Hammarström et al., 2023; Mufwene, 2006; Hall, 1958; Faine, 1937) **relatives** for Haitian, Papiamentu, and Sango. We also experiment with transfer between these **Creole languages**, as previous works have argued that Creole languages have linguistic commonalities with each other (Daval-Markussen and Bakker, 2012). We next select languages by similarity of documented typological features, calculated as **lang2vec** distances (Littell et al., 2017). For data-driven language selection, we determine language relatedness from similarity of NLLB-200 language embeddings (**NLLB representations**; NLLB Team et al., 2022), as well as Subword Evenness (**SuE**; Pelloni et al., 2022). For details of our language selection processes, see Appendix C.

## 2.2 Control experiments

As an experimental control, we identify three language groups with no linguistic relation to the Creole languages in question: **Uralic** languages, **Dravidian** languages, and **CJK** (*i.e.*, Chinese, Japanese, and Korean). None of these groups share phylogeny or significant amounts of typological features with our Creole languages. The latter two also differ in terms of script. And the final group is distinct in that its languages also bear no phylogenetic relation to each other. As a final experimental control, we use a randomly initialized, untrained adapter (**init**). We use this untrained adapter in two ways: as a replacement for the Creole language adapter, and souped with the Creole language adapter. As Chronopoulou et al. (2023a) suggest for souping, we use the same random initialization for all untrained and trained adapters.

## 2.3 Experimental Details

Our experiments follow the same general steps. ① We train all language adapters individually on denoising using monolingual data from MADLAD-400 (Kudugunta et al., 2023) that we preprocess

and clean further.<sup>2</sup> ② We add the source language adapter  $s$  to the encoder and the target language adapter  $t$  to the decoder and freeze all adapter and model parameters. In adapter averaging experiments the `init` or the helper language adapters are souped with the  $s$  or  $t$  language adapters in this step as well. ③ We unfreeze the cross-attention of the decoder and use 10k parallel data from NLLB-OPUS (NLLB Team et al., 2022) to do cross-attention fine-tuning (CA-FT). ④ We evaluate MT from  $s \rightarrow t$  on Flores-200 (NLLB Team et al., 2022) using mean BLEU and chrF scores. (See Table 1 for data composition.)

All experiments are conducted using a 600M-parameter distillation of the original 54B-parameter NLLB-200 model (NLLB Team et al., 2022) which has 12 encoder and 12 decoder layers, 16 attention heads, 1,024 dimensions, and a model vocabulary of over 250,000 sentence-piece tokens shared across all languages. We adopt the adapter training and the hyperparameters from Üstün et al. (2021), training bottleneck adapters on 10,000 training samples for a maximum of 100,000 training steps with early stopping, a batch size of 8, 4,000 warmup steps, and the Adam optimizer with a maximum learning rate of  $2e-4$  with a gradient accumulation of 8 steps. All adapters initialized from the same random seed as Chronopoulou et al. (2023a) show that this is necessary for successful souping. We train CA-FT for 1 epoch with early stopping. We soup adapters with the same weight except when averaging with the `init` adapter where the weight of the Creole language to `init` adapter is 1 : 3, to closer approximate the other setups where the Creole language adapter is souped with three other adapters.

## 3 Results and Analysis

Here we present only Creole-to-English BLEU scores. We conducted preliminary experiments using only 800 aligned fine-tuning segments on these and the inverse translation directions, and saw the same trends shown in our results here. We decided to prioritize into-English MT for our main experiments after seeing small adapter gains across the board for into-Creole. (See Appendix D for our full results suite.) Table 2 shows BLEU improvements of our approaches over both baselines: the base MT model with CA-FT, and the  $s$  and  $t$  language adapters without souping (Üstün et al., 2021). The

<sup>2</sup>See Appendix B for preprocessing details.

Experiment	hat → eng	pap → eng	sag → eng
Base Model (CA-FT)	33.37	38.97	10.89
<i>s</i> and <i>t</i> Adapters	32.33	40.04	11.40
📖 Untrained <i>s</i> Adapter	37.07	45.01	14.91
🌿 IE Transfer	36.44	46.35	12.46
🌿 NC Transfer	36.06	46.69	12.29
❤️ Creole Transfer	35.25	46.23	12.76
📖 lang2vec	36.54	47.04	13.07
👤 NLLB Vec	35.80	46.91	12.80
🗑️ Subword Evenness	36.36	47.03	13.12
Mean Improvement	+2.70	+7.74	+1.86
📖 Untrained Souping	37.42	46.34	13.41
🗑️ Uralic	37.06	47.00	13.58
🗑️ CJK	36.41	47.17	13.33
🗑️ Dravidian	36.55	47.27	13.27
Mean Improvement	+3.49	+7.98	+2.51

Table 2: Mean BLEU scores for Creole to English MT experiments over a diverse set of conditions for selecting transfer languages, as motivated by phylogeny (🌿), language group (❤️), typological databases (📖), model representations (👤), and tokenization (🗑️). Experimental results are also compared against control experiments with unrelated languages (🗑️) and untrained adapters (📖). We highlight mean performance improvements over the Base Model (CA-FT) for both transfer and controls.

improvements, however, are not particularly differentiated between various transfer and control scenarios, and notably, untrained adapters match or exceed performance of trained ones.

**Regularization, Not Transfer?** Two facets of our results indicate that cross-lingual transfer does not explain the success of our methodology. The first is that adapter souping improves performance over the baselines, *regardless* of the criteria used to select transfer languages. The unrelated control languages (*i.e.*, Dravidian, Uralic, and CJK) achieve performance gains on par with languages selected through principled means (*e.g.*, genealogy and typology), without marked differences in BLEU or chrF scores. This holds for both the source and target language adapters in the encoder and decoder, respectively. The second is the commensurate success of the untrained adapters in averaging. Souping adapters with ‘meaningful’ language adapters is no better than souping with random noise, indicating that cross-lingual transfer (or any information transfer) is likely not responsible for performance gains. These findings seem more consistent with previous works on regularization, which find that adding noise can improve a model’s ability to gener-

FT examples:	800	10k
Base Model (CA-FT)	45.45	45.53
<i>s</i> and <i>t</i> Adapters	38.58	41.92
🌿 spa + por + oci Souping	41.87	43.74
📖 Untrained Souping	43.97	44.75

Table 3: BLEU scores for Catalan to English MT.

alize and avoid overfitting (Hochreiter and Schmidhuber, 1994; Poole et al., 2014; Moradi et al., 2020) (which may be pertinent given both our limited train sets and train/test domain mismatch).

We perform multiple analyses to test our hypothesis that adapters are regularizing. First, we compared the use of untrained adapters to the use of adapters of known closely related languages. Because the close relatives of Creole languages are disputed, we turned to another relatively low resource language, Catalan, which has three close relatives also supported by NLLB-200. Spanish, Portuguese, and Occitan are closely related to Catalan by effectively every reasonable linguistic criterion. Table 3 shows different trends for Catalan (a language on which NLLB’s baseline performance is notably high). In this scenario, none of the adapter approaches improve performance over the baseline, once again indicating that the monolingual denoising adapter training on MADLAD data is not positively contributing. And once again, even with these highly attested close linguistic relationships, we see the random adapter comes nearest to recovering the baseline score in both data settings. This set of results underscores the possibility that CA-FT without randomized regularization may lead to overfitting (hence the performance drops). We also experimented souping only one or two related languages for both Catalan→English and Papiamentu→English but saw no improvements.

We observed training loss, gradient norms, and dev loss for CA-FT of the experiments presented in this section. We compare untrained adapter souping (hypothesized regularization) with simple *s* and *t* adapters (unregularized), finding that the unregularized case has moderately higher gradient norms and dev loss, consistent with our hypothesis. The unregularized case also sees higher train loss, which is inconclusive regarding our hypothesis since an overfit model may be stuck in a local minimum or be overfit on a distribution varying from the CA-FT data. (See Appendix E.)

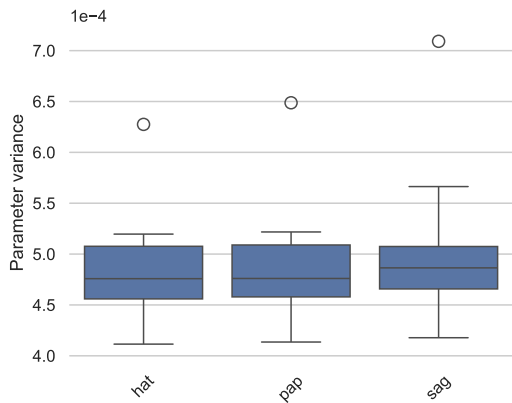


Figure 2: Parameter variance between the pretrained Creole adapters (represented by the outliers) vs the souped adapters of various experimental conditions (shown in the boxplots).

We also show that parameter variance decreases substantially when souping adapters compared to the pretrained Creole adapters (Figure 2), further supporting our regularization explanation.

**Qualitative Evaluation** Beyond computational analyses, we further assess the quality of translations across our experimental conditions. To start, a subset of the authors create a simple annotation guideline, inspired by MQM (Burchardt, 2013), to help identify errors regarding vocabulary, grammar, and other translation issues (*e.g.*, addition or omission). Then, one author, a native Haitian speaker, carries out a linguistically-grounded manual evaluation of Haitian  $\rightarrow$  English translation across the baseline of *s* and *t* Adapters, the IE Transfer, and the Untrained Souping. The evaluation involves 33 samples per condition presented to the annotator in three shuffled sets (*i.e.*, 99 sentences in total) to avoid biasing him. The primary focus falls on grammar and translation issues (For a condensed list of errors, see Table 9 in Appendix F). Although a small sample size, we observe that adapter souping results in less grammatical and lexical errors when applying the pretrained Haitian and English adapters with CA-FT. While both IE Transfer and Untrained Souping are comparable in reducing lexical errors, the latter does better in avoiding grammatical errors. This may indicate a comparative success of regularization when adding random noise that is also reflected in the +0.98 BLEU score Untrained Souping has over IE Transfer for Haitian  $\rightarrow$  English (see Table 2). We believe that even a small manual translation evaluation demonstrates the value of involving native speakers in MT.

## 4 Conclusion

Inspired by prior success of monolingual adapters and CA-FT for low-resource MT, we corroborate prior findings that adapters can help by consistently improving BLEU scores for a set of Creole languages. However, we differ from prior research by finding no evidence of linguistic or informational transfer from adapters, even across a wide array of linguistic relationships. We hypothesize that most adapter aid in this setting is from regularization and find some analytical support for this.

## Limitations

**Scalability** This work can only be applied to a small subset of all Creoles that are present in larger benchmarks (Lent et al., 2024; Robinson et al., 2024) due to the fact that they are low-resource languages – in certain cases, extremely so –, so there is an issue of data scarcity and a limited support by MT models. This data scarcity extends to also the African relatives of the Creoles we target in this paper. Our work is further limited by the limited coverage by typological resources. Grambank (Skirgård et al., 2023) supports only Haitian, not Papiamentu and Sango, and while the lang2vec covers more languages, the coverage is limited and uneven. Despite these constraints, however, we believe our selection of Creoles (*i.e.*, Haitian, Papiamentu, and Sango) are diverse enough to make high-level hypotheses about transfer learning for Creoles; these languages have different relatives (*e.g.*, inheriting large vocabulary from French, Portuguese, and Ngbandi, respectively) and are spoken across different regions (*i.e.*, the Caribbean and Africa). Moreover, our experimentation over these languages is **extensive**, with careful investigation of 6 transfer scenarios, 1 `init` baseline, and 3 principled controls. So while this work cannot scale *widely* across Creoles, we investigate transfer scenarios for each Creole *deeply*.

**Domain of Data** Similar to many other low-resource languages, much of the available data for Creoles comes from the religious domain (*e.g.*, Bible translations), which has a characteristic style that is typically not representative of language use in other contexts (Mielke et al., 2019). Moreover, religious domain content may not necessarily be culturally appropriate for some communities (Hershovich et al., 2022; Mager et al., 2023). In order to be truly useful for end users, data sourced from a

wide variety of domains will ultimately be a necessary component for the development of robust MT systems for Creoles. Until more such data becomes available, we believe that work at the intersection of transfer learning and domain adaptation could be greatly beneficial for Creole NLP.

### Difficulty in Verifying Neural Phenomena

While we find some analytical evidence that our adapters are boosting performance by regularizing fine-tuning, it is difficult to verify this with certainty. Regularization is a mathematical phenomenon that can be demonstrated rigorously, but such involved analysis is outside the scope of this current work; hence we leave these important analytical details to future work.

### Ethics Statement

As Creole languages are often marginalized, we recognize the necessity of heightened ethical scrutiny for research involving these languages. We thus engage directly with considerations from [Lent et al. \(2022b\)](#), [Mager et al. \(2023\)](#), and [Bird \(2020\)](#), to ensure that our work is in accordance with the needs and desires of speakers. To this end, we note that speakers of many different Creole languages actively *want* and/or *need* improved MT technology for their language. In the context of Haitian Creole, for example, MT technology is highly desirable for facilitating interactions with non-Haitian speakers (e.g., the neighboring, Spanish-speaking Dominican Republic, and foreign workers at NGOs operating in Haiti). Additionally, MT has the potential to greatly improve access to information for Haitians. [DeGraff et al. \(2022\)](#) argues that the lack of access to education in Creole is a critical barrier to social and economic mobility in Haiti. Accordingly, reliable MT for Haitian Creole could make available high-quality educational resources from platforms such as edX, Coursera, MITx, and Udemy, for example. Thus, if MT systems could offer accurate translations into Haitian Creole, they would become transformative tools for advancing the country’s development, breaking down educational barriers, and fostering greater social equity. We note similar potential for MT for speakers of Sango in Central African Republic, Chad, and Democratic Republic of the Congo and ([African Development Bank Group, 2024](#)). And while the socioeconomic context of Aruba, Bonaire, and Curaçao is quite different from that of aforementioned Haiti or Central African Republic, speakers’ desire

for language technology for Papiamentu is well documented through projects like [papiamentu.ai](#).

### Acknowledgments

MF, JB, and HL were supported by the Carlsberg Foundation, under the *Semper Ardens: Accelerate* program (project nr. CF210454); the same project also supported the annotation work performed by DJB. EL was funded by the Google Award for Inclusion Research program (awarded to JB and HL for the “CREOLE: Creating Resources for Disadvantaged Language Communities” project). We extend our thanks to Mike Zhang, Esther Ploeger, and the other members of AAU-NLP, for their helpful feedback on our manuscript.

### References

- David Adelani, Jesujoba Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruitter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajuddeen Gwadabe, Freshia Sackey, Bonaventure F. P. Dossou, Chris Emezue, Colin Leong, Michael Beukman, Shamsuddeen Muhammad, Guyo Jarso, Oreen Yousuf, Andre Niyongabo Rubungo, Gilles Hacheme, Eric Peter Wairagala, Muhammad Umair Nasir, Benjamin Ajibade, Tunde Ajayi, Yvonne Gitau, Jade Abbott, Mohamed Ahmed, Millicent Ochieng, Anuoluwapo Aremu, Perez Ogayo, Jonathan Mukiibi, Fatoumata Ouoba Kabore, Godson Kalipe, Derguene Mbaye, Allahsera Auguste Tapo, Victoire Memdjokam Koagne, Edwin Munkoh-Buabeng, Valencia Wagner, Idris Abdulmumin, Ayodele Awokoya, Happy Buzaaba, Blessing Sibanda, Andiswa Bukula, and Sam Manthalu. 2022. [A few thousand translations go a long way! leveraging pre-trained models for African news translation](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3053–3070, Seattle, United States. Association for Computational Linguistics.
- David Ifeoluwa Adelani, Jade Abbott, Graham Neubig, Daniel D’souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, Stephen Mayhew, Israel Abebe Azime, Shamsuddeen H. Muhammad, Chris Chinenye Emezue, Joyce Nakatumba-Nabende, Perez Ogayo, Aremu Anuoluwapo, Catherine Gitau, Derguene Mbaye, Jesujoba Alabi, Seid Muhie Yimam, Tajuddeen Rabiou Gwadabe, Ignatius Ezeani, Rubungo Andre Niyongabo, Jonathan Mukiibi, Verah Otiende, Iroro Orife, Davis David, Samba Ngom, Tosin Adewumi, Paul Rayson, Mofetoluwa Adeyemi, Gerald Muriuki, Emmanuel Anebi, Chiamaka Chukwunke, Nkiruka Odu, Eric Peter Wairagala, Samuel Oyerinde, Clemencia Siro, Tobius Saul Bateesa, Temilola Oloyede, Yvonne Wambui, Victor Akin-

- ode, Deborah Nabagereka, Maurice Katusiime, Ayodele Awokoya, Mouhamadane MBOUP, Dibora Gebreyohannes, Henok Tilaye, Kelechi Nwaike, De-gaga Wolde, Abdoulaye Faye, Blessing Sibanda, Orevaoghene Ahia, Bonaventure F. P. Dossou, Kelechi Ogueji, Thierno Ibrahima DIOP, Abdoulaye Diallo, Adewale Akinfaderin, Tendai Marengereke, and Salomey Osei. 2021. [MasakhaNER: Named entity recognition for African languages](#). *Transactions of the Association for Computational Linguistics*, 9:1116–1131.
- African Development Bank Group. 2024. [Central african republic: New digital era dawns with debut of high-speed internet cable](#). *African Development Bank Group*.
- Hamidreza Amirzadeh, Sadegh Jafari, Anika Harju, and Rob van der Goot. 2025. [data2lang2vec: Data driven typological features completion](#). In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 6520–6529, Abu Dhabi, UAE. Association for Computational Linguistics.
- Ruth-Ann Armstrong, John Hewitt, and Christopher Manning. 2022. [JamPatoisNLI: A jamaican patois natural language inference dataset](#). In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 5307–5320, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Steven Bird. 2020. [Decolonising speech and language technology](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3504–3519, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Aljoscha Burchardt. 2013. [Multidimensional quality metrics: a flexible system for assessing translation quality](#). In *Proceedings of Translating and the Computer 35*, London, UK. Aslib.
- Chris Callison-Burch, Philipp Koehn, Christof Monz, and Omar Zaidan. 2011. [Findings of the 2011 workshop on statistical machine translation](#). In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 22–64, Edinburgh, Scotland. Association for Computational Linguistics.
- Bernard Caron, Marine Courtin, Kim Gerdes, and Sylvain Kahane. 2019. [A surface-syntactic UD treebank for Naija](#). In *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019)*, pages 13–24, Paris, France. Association for Computational Linguistics.
- Alexandra Chronopoulou, Matthew Peters, Alexander Fraser, and Jesse Dodge. 2023a. [AdapterSoup: Weight averaging to improve generalization of pretrained language models](#). In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 2054–2063, Dubrovnik, Croatia. Association for Computational Linguistics.
- Alexandra Chronopoulou, Dario Stojanovski, and Alexander Fraser. 2023b. [Language-family adapters for low-resource multilingual neural machine translation](#). In *Proceedings of the Sixth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2023)*, pages 59–72, Dubrovnik, Croatia. Association for Computational Linguistics.
- Christopher Clarke, Roland Daynauth, Jason Mars, Charlene Wilkinson, and Hubert Devonish. 2024. [GuyLingo: The Republic of Guyana creole corpora](#). In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 792–798, Mexico City, Mexico. Association for Computational Linguistics.
- Raj Dabre and Aneerav Sukhoo. 2022. [Kreol-MorisienMT: A dataset for mauritian creole machine translation](#). In *Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022*, pages 22–29, Online only. Association for Computational Linguistics.
- Aymeric Daval-Markussen and Peter Bakker. 2012. [Explorations in creole research with phylogenetic tools](#). In *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*, pages 89–97, Avignon, France. Association for Computational Linguistics.
- Wietse de Vries, Martijn Wieling, and Malvina Nissim. 2022. [Make the best of cross-lingual transfer: Evidence from POS tagging with over 100 languages](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7676–7685, Dublin, Ireland. Association for Computational Linguistics.
- Michel DeGraff, William Scott Frager, and Haynes Miller. 2022. [Language policy in haitian education](#). *Journal of Haitian Studies*, 28(2):33–95.
- Daryna Dementieva, Valeriia Khylenko, and Georg Groh. 2025. [Cross-lingual text classification transfer: The case of Ukrainian](#). In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 1451–1464, Abu Dhabi, UAE. Association for Computational Linguistics.
- Matthew S. Dryer and Martin Haspelmath, editors. 2013. [WALS Online \(v2020.3\)](#). Zenodo.
- Jules Faine. 1937. [Philologie creole: Études historique et étymologiques sur la langue creoles d’Haïti](#). Imprimerie de l’Etat, Port-au-Prince.
- Fahim Faisal and Antonios Anastasopoulos. 2022. [Phylogeny-inspired adaptation of multilingual models to new languages](#). In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 434–452, Online only. Association for Computational Linguistics.

- Fahim Faisal and Antonios Anastasopoulos. 2024. [Data-augmentation-based dialectal adaptation for LLMs](#). In [Proceedings of the Eleventh Workshop on NLP for Similar Languages, Varieties, and Dialects \(VarDial 2024\)](#), pages 197–208, Mexico City, Mexico. Association for Computational Linguistics.
- Robert E. Frederking, Ralf D. Brown, and Christopher Hogan. 1997. [The DIPLOMAT rapid development speech MT system](#). In [Proceedings of Machine Translation Summit VI: Systems](#), pages 261–262, San Diego, California.
- Robert Hall. 1958. Creole languages and genetic relationships. [Word](#), 14:367–373.
- Harald Hammarström, Sebastian Nordhoff, Robert Forkel, Martin Haspelmath, and Sebastian Bank. 2023. [Glottolog](#). Online database.
- Daniel Hershcovich, Stella Frank, Heather Lent, Miryam de Lhoneux, Mostafa Abdou, Stephanie Brandl, Emanuele Bugliarello, Laura Cabello Piqueras, Ilias Chalkidis, Ruixiang Cui, Constanza Fierro, Katerina Margatina, Phillip Rust, and Anders Søgaard. 2022. [Challenges and strategies in cross-cultural NLP](#). In [Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics \(Volume 1: Long Papers\)](#), pages 6997–7013, Dublin, Ireland. Association for Computational Linguistics.
- Sepp Hochreiter and Jürgen Schmidhuber. 1994. [Simplifying neural nets by discovering flat minima](#). In [Advances in Neural Information Processing Systems](#), volume 7. MIT Press.
- Neil Houlsby, Andrei Giurgiu, Stanisław Jastrzebski, and Bruna Morrone. 2019. [Parameter-Efficient Transfer Learning for NLP](#). In [Proceedings of the 36th International Conference on Machine Learning](#), page 10, Long Beach, California.
- Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. [LoRA: Low-Rank Adaptation of Large Language Models](#). [arXiv preprint](#). ArXiv:2106.09685 [cs].
- Silvia Kouwenberg and John Victor Singler. 2009. [The Handbook of Pidgin and Creole Studies](#). John Wiley & Sons.
- Sneha Kudugunta, Isaac Caswell, Biao Zhang, Xavier Garcia, Christopher A. Choquette-Choo, Katherine Lee, Derrick Xin, Aditya Kusupati, Romi Stella, Ankur Bapna, and Orhan Firat. 2023. [Madlad-400: A multilingual and document-level large audited dataset](#). [Preprint](#), arXiv:2309.04662.
- Claire Lefebvre. 2011. Substrate features in the properties of verbs in three atlantic creoles: Haitian Creole, Saramaccan and Papiamentu. In Claire Lefebvre, editor, [Creoles, their Substrates, and Language Typology](#), pages 127–154. John Benjamins, Amsterdam.
- Heather Lent, Emanuele Bugliarello, and Anders Søgaard. 2022a. [Ancestor-to-creole transfer is not a walk in the park](#). In [Proceedings of the Third Workshop on Insights from Negative Results in NLP](#), pages 68–74, Dublin, Ireland. Association for Computational Linguistics.
- Heather Lent, Kelechi Ogueji, Miryam de Lhoneux, Orevaoghene Ahia, and Anders Søgaard. 2022b. [What a creole wants, what a creole needs](#). In [Proceedings of the Thirteenth Language Resources and Evaluation Conference](#), pages 6439–6449, Marseille, France. European Language Resources Association.
- Heather Lent, Kushal Tatariya, Raj Dabre, Yiyi Chen, Marcell Fekete, Esther Ploeger, Li Zhou, Ruth-Ann Armstrong, Abee Eijansantos, Catriona Malau, Hans Erik Heje, Ernests Lavrinovics, Diptesh Kanojia, Paul Belony, Marcel Bollmann, Loïc Grobol, Miryam de Lhoneux, Daniel Hershcovich, Michel DeGraff, Anders Søgaard, and Johannes Bjerva. 2024. [CreoleVal: Multilingual multitask benchmarks for creoles](#). [Transactions of the Association for Computational Linguistics](#), 12:950–978.
- Patrick Littell, David R. Mortensen, Ke Lin, Katherine Kairis, Carlisle Turner, and Lori Levin. 2017. [URIEL and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors](#). In [Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers](#), pages 8–14, Valencia, Spain. Association for Computational Linguistics.
- Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohata, Tenghao Huang, Mohit Bansal, and Colin Raffel. 2022a. [Few-Shot Parameter-Efficient Fine-Tuning is Better and Cheaper than In-Context Learning](#). [arXiv preprint](#). ArXiv:2205.05638 [cs].
- Zhengyuan Liu, Shikang Ni, Ai Ti Aw, and Nancy F. Chen. 2022b. [Singlish message paraphrasing: A joint task of creole translation and text normalization](#). In [Proceedings of the 29th International Conference on Computational Linguistics](#), pages 3924–3936, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.
- Manuel Mager, Elisabeth Mager, Katharina Kann, and Ngoc Thang Vu. 2023. [Ethical considerations for machine translation of indigenous languages: Giving a voice to the speakers](#). In [Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics \(Volume 1: Long Papers\)](#), pages 4871–4897, Toronto, Canada. Association for Computational Linguistics.
- Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus Huber, editors. 2013. [APiCS Online](#). Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Sabrina J. Mielke, Ryan Cotterell, Kyle Gorman, Brian Roark, and Jason Eisner. 2019. [What kind of language is hard to language-model?](#) In [Proceedings](#)



- of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4975–4989, Florence, Italy. Association for Computational Linguistics.
- 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, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.
- Reza Moradi, Reza Berangi, and Behrouz Minaei. 2020. [A survey of regularization strategies for deep models](#). [Artificial Intelligence Review](#), 53(6):3947–3986.
- Salikoko Mufwene. 2006. Pidgins and creoles. In Braj B. Kachru, Yamuna Kachru, and Cecil L. Nelson, editors, [The Handbook of World Englishes](#), pages 313–327. Blackwell, Hoboken, NJ.
- Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Seid Muhie Yimam, David Ifeoluwa Adelan, Ibrahim Said Ahmad, Nedjma Ousidhoum, Abinew Ali Ayele, Saif Mohammad, Meriem Beloucif, and Sebastian Ruder. 2023. [SemEval-2023 task 12: Sentiment analysis for African languages \(AfriSenti-SemEval\)](#). In [Proceedings of the 17th International Workshop on Semantic Evaluation \(SemEval-2023\)](#), pages 2319–2337, Toronto, Canada. Association for Computational Linguistics.
- Shamsuddeen Hassan Muhammad, David Ifeoluwa Adelan, Sebastian Ruder, Ibrahim Sa'id Ahmad, Idris Abdulmumin, Bello Shehu Bello, Monojit Choudhury, Chris Chinenye Emezue, Saheed Salahudeen Abdullahi, Anuoluwapo Aremu, Alípio Jorge, and Pavel Brazdil. 2022. [NaijaSenti: A Nigerian Twitter sentiment corpus for multilingual sentiment analysis](#). In [Proceedings of the Thirteenth Language Resources and Evaluation Conference](#), pages 590–602, Marseille, France. European Language Resources Association.
- Antonio Carvalho Neto, Fernanda Versiani, Kelly Pelizari, Carolina Mota-Santos, and Gustavo Abreu. 2020. Latin american, african and asian immigrants working in brazilian organizations: facing the language barrier. [Revista Economia & Gestão](#), 20(55).
- NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia-Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. 2022. No language left behind: Scaling human-centered machine translation.
- Kelechi Ogueji and Orevaoghene Ahia. 2019. [Pidginunmt: Unsupervised neural machine translation from west african pidgin to english](#). [arXiv preprint arXiv:1912.03444](#).
- Wuraola Fisayo Oyewusi, Olubayo Adekanmbi, and Olalekan Akinsande. 2020. [Semantic enrichment of nigerian pidgin english for contextual sentiment classification](#). [arXiv preprint arXiv:2003.12450](#).
- Marinela Parovic, Alan Ansell, Ivan Vulić, and Anna Korhonen. 2023. [Cross-lingual transfer with target language-ready task adapters](#). In [Findings of the Association for Computational Linguistics: ACL 2023](#), pages 176–193, Toronto, Canada. Association for Computational Linguistics.
- Marinela Parović, Goran Glavaš, Ivan Vulić, and Anna Korhonen. 2022. [BAD-X: Bilingual adapters improve zero-shot cross-lingual transfer](#). In [Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies](#), pages 1791–1799, Seattle, United States. Association for Computational Linguistics.
- Olga Pelloni, Anastassia Shaitarova, and Tanja Samardzic. 2022. [Subword evenness \(SuE\) as a predictor of cross-lingual transfer to low-resource languages](#). In [Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing](#), pages 7428–7445, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Jonas Pfeiffer, Sebastian Ruder, Ivan Vulić, and Edoardo Maria Ponti. 2024. [Modular deep learning](#). Preprint, arXiv:2302.11529.
- Jonas Pfeiffer, Ivan Vulić, Iryna Gurevych, and Sebastian Ruder. 2020. [MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer](#). In [Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing \(EMNLP\)](#), pages 7654–7673, Online. Association for Computational Linguistics.
- Jonas Pfeiffer, Ivan Vulić, Iryna Gurevych, and Sebastian Ruder. 2021. [UNKs everywhere: Adapting multilingual language models to new scripts](#). In [Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing](#), pages 10186–10203, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- 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, Florence, Italy. Association for Computational Linguistics.

- Edoardo Maria Ponti, Helen O’Horan, Yevgeni Berzak, Ivan Vulić, Roi Reichart, Thierry Poibeau, Ekaterina Shutova, and Anna Korhonen. 2019. [Modeling language variation and universals: A survey on typological linguistics for natural language processing](#). *Computational Linguistics*, 45(3):559–601.
- Ben Poole, Jascha Sohl-Dickstein, and Surya Ganguli. 2014. [Analyzing noise in autoencoders and deep networks](#). *CoRR*, abs/1406.1831.
- Vipul Rathore, Rajdeep Dhingra, Parag Singla, and Mausam. 2023. [ZGUL: Zero-shot generalization to unseen languages using multi-source ensembling of language adapters](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6969–6987, Singapore. Association for Computational Linguistics.
- John R Rickford and John McWhorter. 1997. Language contact and language generation: Pidgins and creoles. In Florian Coulmas, editor, *The handbook of sociolinguistics*, pages 238–256. Blackwell, Malden, MA.
- Nathaniel Robinson, Raj Dabre, Ammon Shurtz, Rasul Dent, Onenamiyi Onesi, Claire Monroc, Loïc Grobol, Hasan Muhammad, Ashi Garg, Naome Etori, Vijay Murari Tiyyala, Olanrewaju Samuel, Matthew Stutzman, Bismarck Odoom, Sanjeev Khudanpur, Stephen Richardson, and Kenton Murray. 2024. [Kreyòl-MT: Building MT for Latin American, Caribbean and colonial African creole languages](#). In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 3083–3110, Mexico City, Mexico. Association for Computational Linguistics.
- Nathaniel Robinson, Cameron Hogan, Nancy Fulda, and David R. Mortensen. 2022. [Data-adaptive transfer learning for translation: A case study in Haitian and jamaican](#). In *Proceedings of the Fifth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2022)*, pages 35–42, Gyeongju, Republic of Korea. Association for Computational Linguistics.
- Nathaniel Romney Robinson, Matthew Dean Stutzman, Stephen D. Richardson, and David R Mortensen. 2023. [African substrates rather than european lexifiers to augment african-diaspora creole translation](#). In *4th Workshop on African Natural Language Processing*.
- Hedvig Skirgård, Hannah J. Haynie, Damián E. Blasi, Harald Hammarström, Jeremy Collins, Jay J. Latache, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Sam Passmore, Angela Chira, Luke Maurits, Russell Dinnage, Michael Dunn, Ger Reesink, Ruth Singer, Claire Bower, Patience Epps, Jane Hill, Outi Vesakoski, Martine Robbeets, Noor Karolin Abbas, Daniel Auer, Nancy A. Bakker, Giulia Barbos, Robert D. Borges, Swintha Danielsen, Luise Dorenbusch, Ella Dorn, John Elliott, Giada Falcone, Jana Fischer, Yustinus Ghanggo Ate, Hannah Gibson, Hans-Philipp Göbel, Jemima A. Goodall, Victoria Gruner, Andrew Harvey, Rebekah Hayes, Leonard Heer, Roberto E. Herrera Miranda, Nataliia Hübler, Biu Huntington-Rainey, Jessica K. Ivani, Marilen Johns, Erika Just, Eri Kashima, Carolina Kipf, Janina V. Klingenberg, Nikita König, Aikaterina Koti, Richard G. A. Kowalik, Olga Krasnoukhova, Nora L.M. Lindvall, Mandy Lorenzen, Hannah Lutzenberger, Tônia R.A. Martins, Celia Mata German, Suzanne van der Meer, Jaime Montoya Samamé, Michael Müller, Saliha Muradođlu, Kelsey Neely, Johanna Nickel, Miina Norvik, Cheryl Akinyi Oluoch, Jesse Peacock, India O.C. Pearey, Naomi Peck, Stephanie Petit, Sören Pieper, Mariana Poblete, Daniel Prestipino, Linda Raabe, Amna Raja, Janis Reimringer, Sydney C. Rey, Julia Rizaew, Eloisa Ruppert, Kim K. Salmon, Jill Sammet, Rhiannon Schembri, Lars Schlabach, Frederick W.P. Schmidt, Amalia Skilton, Wikaliler Daniel Smith, Hilário de Sousa, Kristin Sverredal, Daniel Valle, Javier Vera, Judith Voß, Tim Witte, Henry Wu, Stephanie Yam, Jingting Ye, Maisie Yong, Tessa Yuditha, Roberto Zariquiey, Robert Forkel, Nicholas Evans, Stephen C. Levinson, Martin Haspelmath, Simon J. Greenhill, Quentin D. Atkinson, and Russell D. Gray. 2023. [Grambank reveals global patterns in the structural diversity of the world’s languages](#). *Science Advances*, 9.
- Maali Tars, Andre Tättar, and Mark Fišel. 2021. [Extremely low-resource machine translation for closely related languages](#). In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 41–52, Reykjavik, Iceland (Online). Linköping University Electronic Press, Sweden.
- Ahmet Üstün, Alexandre Berard, Laurent Besacier, and Matthias Gallé. 2021. [Multilingual unsupervised neural machine translation with denoising adapters](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6650–6662, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Ahmet Üstün, Arianna Bisazza, Gosse Bouma, Gertjan van Noord, and Sebastian Ruder. 2022. [Hyper-X: A unified hypernetwork for multi-task multilingual transfer](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7934–7949, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Raúl Vázquez, Timothee Mickus, Jörg Tiedemann, Ivan Vulić, and Ahmet Üstün, editors. 2024. [Proceedings of the 1st Workshop on Modular and Open Multilingual NLP \(MOOMIN 2024\)](#). Association for Computational Linguistics, St Julians, Malta.
- 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 (Volume 1: Long Papers), pages 1732–1744, Vancouver, Canada. Association for Computational Linguistics.

Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, and Ludwig Schmidt. 2022. [Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time](#). In [Proceedings of the 39th International Conference on Machine Learning](#), volume 162 of [Proceedings of Machine Learning Research](#), pages 23965–23998. PMLR.

Fei Yuan, Yinquan Lu, Wenhao Zhu, Lingpeng Kong, Lei Li, Yu Qiao, and Jingjing Xu. 2023. [Lego-MT: Learning detachable models for massively multilingual machine translation](#). In [Findings of the Association for Computational Linguistics: ACL 2023](#), pages 11518–11533, Toronto, Canada. Association for Computational Linguistics.

Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. [Transfer learning for low-resource neural machine translation](#). In [Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing](#), pages 1568–1575, Austin, Texas. Association for Computational Linguistics.

Just Zwennicker and David Stap. 2022. Towards a general purpose machine translation system for sranantongo. [arXiv preprint arXiv:2212.06383](#).

## A Literature Review

NLP research for Creole languages is still in relatively early stages. Some large-scale multilingual MT systems have limited Creole language support: NLLB (NLLB Team et al., 2022) supports five Creole languages, and Google Translate<sup>3</sup> and Lego-MT (Yuan et al., 2023) each support eight. Other works have focused specifically on building MT for Creole languages Haitian (Robinson et al., 2022; Callison-Burch et al., 2011; Frederking et al., 1997), Singlish (Wang et al., 2017; Liu et al., 2022b), Nigerian Pidgin (Adelani et al., 2022; Ogueji and Ahia, 2019), Mauritian Creole (Dabre and Sukhoo, 2022), Sranan Tongo (Zwennicker and Stap, 2022), Guyanese Creole (Clarke et al., 2024), and Jamaican Patois (Robinson et al., 2022). Some others have focused on building Creole language-specific NLP tools outside of MT for Nigerian Pidgin (Muhammad et al., 2023, 2022; Adelani et al., 2021; Oyewusi et al., 2020; Caron et al., 2019), Jamaican Patois (Armstrong et al., 2022), and Martinican Creole (Mompelat et al., 2022).

<sup>3</sup><https://translate.google.com/> as of 8 Feb., 2025

More recently, researchers have developed large scale multi-Creole-language NLP applications. Lent et al. (2024) developed CreoleVal, a multitask benchmark for Creoles including an MT dataset and model for 28 Creole languages. Robinson et al. (2024) expanded on Lent et al.’s (2024) work, building genre-diverse MT datasets and models for 41 Creole languages of the Caribbean, Latin America, and Colonial Africa. Perhaps most pertinent to our current work are past studies into cross-lingual transfer to benefit Creole NLP. In tightly controlled experiments, Robinson et al. (2023) found indications that Haitian and Jamaican Patois’ Niger-Congo relatives may be more helpful as transfer languages than their Indo-European relatives, for MT applications. Lent et al. (2022a), however, did not identify clear advantages from a language’s relatedness to a Creole language in terms of its usefulness in transfer learning. Given the sparse research and even sparser positive results in the area, optimal transfer learning for Creole NLP remains an open question.

## B Data Preprocessing

The MadLad corpus is organized in series of internet-scraped documents, which the MadLad team cross-checks for noise (Kudugunta et al., 2023). For training the monolingual adapters, we use the *MadLad-Clean* subsplit and we perform further preprocessing, since we noted after manual inspection that noise can still be present in the documents. Firstly, the MadLad documents are split into sentences using the SentenceSplitter<sup>4</sup> library.

For each sentence in the document, we define the following rules, if any of the rules is breached, then the given sentence is not included in the final monolingual training corpus.

1. Sentence ends with a terminal punctuation mark (’,’, ’!’, ’ ” ’, ’?’).
2. Ratio of numerical characters with respect to all other characters is less than **0.25**.
3. Sentence starting with a bullet-point.
4. Sentence containing a *lorem ipsum* substring.
5. Ratio of stop-words with respect to all other tokens is less than **0.6** using whitespace tokenization.
6. Ratio of token count that does not contain alpha characters with respect to total token count is less than **0.25**, using whitespace tokenization.

<sup>4</sup><https://github.com/mediacloud/sentence-splitter>

7. Sentence length is less than **14** and more than **300** characters.
8. Ratio of capitalized characters is less than **0.2**.
9. Ratio of alpha characters with respect to all other characters is more than **0.6**.
10. Language identification does not match the expected language. For this we employ FastText model <sup>5</sup>.

This preprocessing removed approximately 20-23% of sentences, depending on the language. The same preprocessing steps were also applied to NLLB parallel data used for cross-attention fine-tuning, where preprocessing removed approximately 30-50% of the sentences.

## C Selecting Languages for Transfer

Table 4 displays the different transfer languages we selected in our experiments. Below we detail our processes of selecting them.

### C.1 Phylogeny-based Transfer

Faisal and Anastasopoulos (2022) followed principles of language phylogenetics to inform their transfer learning experiments, training adapters on combined Indo-European data, and for subfamilies of Indo-European such as Germanic, Romance, Indo-Aryan, etc. To translate for French, for example, they stacked a French adapter on top of a Romance and then an Indo-European adapter at inference time. Creole phylogeny is heavily disputed, however (Kouwenberg and Singler, 2009). Linguists have debated whether Atlantic Creole languages such as Haitian and Papiamentu are best understood as Niger-Congo (NC) languages with extensive Indo-European (IE) lexical borrowings (Lefebvre, 2011; Rickford and McWhorter, 1997), or as IE languages that developed with strong NC influence (Faine, 1937; Hall, 1958; Hammarström et al., 2023; Mufwene, 2006). Considering this lack of consensus, we select two groups of related languages for each Creole language: IE relatives and NC relatives. In this work, we made a selection based on APICS (Michaelis et al., 2013) and linguistic references.

### C.2 Typological Feature-based Transfer

Information about the typological features of languages is typically documented in databases such as APICS (Michaelis et al., 2013), WALS (Dryer

<sup>5</sup><https://huggingface.co/facebook/fasttext-language-identification>

Condition	Language
<b>Haitian</b>	
Creole Language	hat
IE Relative	fr, es, en
NC Relative	ig, ee, fon
lang2vec	grc, oc, fo
NLLB	pap, epo, ltz
<b>Papiamentu</b>	
Creole Language	pap
IE Relative	nl, pt, es
NC Relative	kg, ak, fon
lang2vec	suz, sr, ug
NLLB	hat, war, tgl
<b>Sango</b>	
Creole Language	sag
IE Relative	fr
NC Relative	kbp
lang2vec	ig, yo, ca
NLLB	kg, tsn, lin
Subword Evenness	kat, rus, zsm
<b>Controls</b>	
Uralic	fin, est, hun
CJK	zho, jap, kor
Dravidian	te, ta, ka
<b>Other Settings</b>	
<b>Hungarian</b>	
lang2vec	fin, tur, als
NLLB	fin, bul, hrv
<b>Yoruba</b>	
lang2vec	tha, ibo, fon
NLLB	ibo, hau, xho

Table 4: All languages used for the various experimental settings.

and Haspelmath, 2013), and Grambank (Skirg ard et al., 2023). The lang2vec library (Littell et al., 2017) assembles typological feature vectors from different sources, enabling measurement of holistic typological distance between languages. The drawbacks of lang2vec include the lack of equal coverage of languages in terms of the number of features, and the coarseness of the linguistic properties (Ponti et al., 2019). To supplant the lack of coverage, we use data2lang2vec which uses textual data to inform feature-prediction of the missing features (Amirzadeh et al., 2025).

### C.3 Data-driven Language Selection Methods

**NLLB Language Embeddings** Typological similarity between languages can also be derived in a data-driven manner, *e.g.*, from language embeddings. To derive the similarity between languages this way, we utilize the language embeddings of the distilled NLLB-200 model which implicitly contain much information with regards to the various languages especially with regards to morphosyntax and vocabulary.

**Subword Evenness** Introduced by Pelloni et al. (2022), Subword Evenness (SuE) is a metric quantifying the differences in the length of subword units for each individual language. It presents another data-driven approach to transfer language selection. In this metric, lower SuE scores have been demonstrated as a good predictor of transfer success to various low-resource languages. Thus, we select the 5 languages in the MADLAD dataset with the lowest SuE angles as calculated when using a BPE tokenizer, see Table 5.

Language Code	Language Code	Script	SuE Angle
-azb	South Azerbaijani	Arabic	66.72
kat	Georgian	Georgian	72.57
zsm	Malay	Latin	79.80
-kas	Kashmiri	Devanagari	80.79
rus	Russian	Cyrillic	81.21

Table 5: MADLAD Languages with the lowest SuE angles calculated using a BPE tokenizer. We note that MADLAD only has Norther Azerbaijani (Latin script), and not South Azerbaijani (Arabic script). Similarly, Kashmiri is only in MADLAD with the Arabic script, not with Devanagari.

### C.4 Control Experiments with Unrelated Languages

To determine whether improvements over Creoles are truly the result of successful cross-lingual trans-

fer, we compare the previously described methods against an equivalent adapter approach using entirely unrelated languages. We use three different sets of such unrelated languages. The first set is made up of the **Uralic languages** Finnish, Estonian, and Hungarian, which are written with the Latin script but have no phylogenetic relation to Haitian, Papiamentu, and Sango. The second set consists of Chinese, Japanese, and Korean – so-called **CJK languages** –, which belong to three distinct language families and are written with three distinct scripts having little to no overlap with Atlantic and African Creole languages. The third group consists of the **Dravidian languages** Tamil, Telugu, and Kannada, which are closely related to each other but also written with three distinct scripts, and again are unrelated to the Creole languages in terms of phylogeny.

## D Full Results

See Table 6 for our results translating between English and Creole languages using only 800 fine-tuning parallel segments.

## E Training Loss

Table 7 compares loss when evaluating on a small subset of the NLLB CA-FT training data. Figure 3 indicates that the adapters we train learn successfully from the MADLAD-400 data on a selection from all language adapters.

## F Native Speaker Evaluation

See Table 9 for a condensed categorization of translation errors under the Untrained Souping (🇵🇸), the IE Transfer (🌱), and the *s* and *t* Adapters (*s* + *t*) conditions.

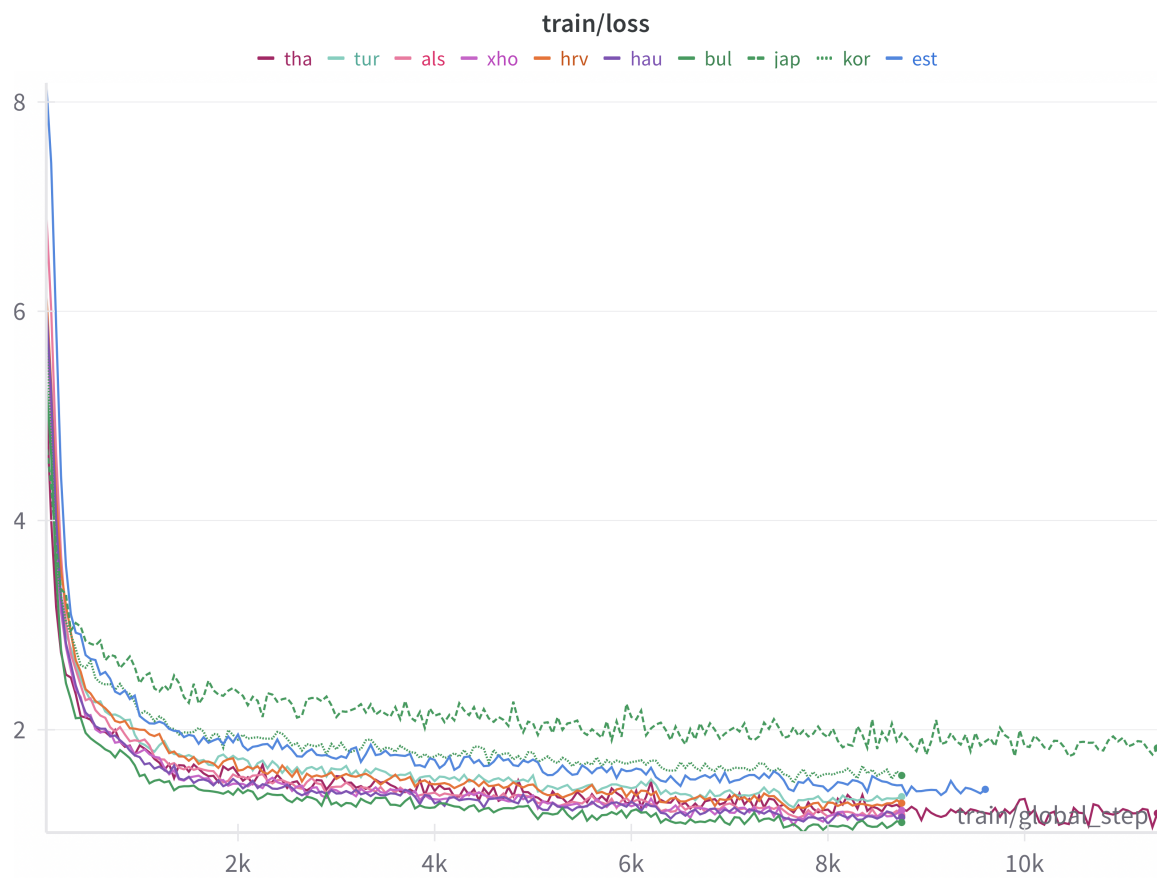


Figure 3: Training loss curves for a subset of adapter training experiments.

Experiments	en → hat	en → pap	en → sag	hat → en	pap → en	sag → en
Baselines						
dNLLB-200 w/ CA-FT	23.59	23.56	11.16	33.36	39.05	10.62
Language Adapters						
Source + Target LA	23.55	26.82	11.90	31.18	38.23	12.19
Phylogeny-Inspired Adaptation						
Target + IE Relatives	24.01	29.35	11.94	35.24	45.48	13.51
Target + NC Relatives	23.98	29.71	11.93	34.96	46.13	13.03
Typological Adaptation						
lang2vec	<b>24.28</b>	<b>30.21</b>	11.71	<b>35.63</b>	<b>46.95</b>	13.79
Target + Data-Driven (NLLB)	24.09	30.17	<b>12.01</b>	34.75	46.09	13.48
Miscellaneous						
Creole Transfer	23.99	29.87	11.71	34.24	44.75	13.48
Subword Evenness	24.23	29.73	11.80	35.20	46.20	<b>14.11</b>
Largest Improvement from dNLLB	+0.69	+6.65	+0.85	+2.27	+7.90	+3.49
Largest Improvement from Üstün	+0.73	+3.39	+0.11	+4.45	+8.63	+1.92
Control						
Uralic	24.31	30.08	11.44	<b>36.01</b>	46.36	<b>15.03</b>
CJK	23.95	29.18	11.60	35.64	46.91	14.71
Dravidian	24.21	29.17	11.69	35.51	46.66	14.45
Random (same init) (1 : 1)	24.21	28.89	11.86	36.07	46.03	14.79
Random (same init) (1 : 3)	24.10	29.65	11.55	37.10	46.66	15.39
Only random (same init)	-	-	-	36.84	45.01	14.90

Table 6: Mean BLEU scores on Flores-200 (after 1 run). Results follow adapter training defaults used in the Üstün et al. (2021) paper. Encoder and decoder adapters are trained together and trained further using cross-attention fine-tuning (CA-FT), and helper and target language adapters are averaged (souped) together. **These scores represent preliminary results, where significantly less data (800 segments) were used to do CA-FT, and thus these results do not correspond to Table 2.**

Experiment	hat → eng	pap → eng	sag → eng
<i>s</i> and <i>t</i> Adapters	0.594	1.106	1.908
Untrained <i>s</i> Adapter	<b>0.476</b>	<b>0.917</b>	<b>1.796</b>

Table 7: Validation loss when evaluating on the CA-FT training data. Averaging with the weights with the same initialization results in lower loss values.

Experiments	en → hu	en → yo	hu → en	yo → en
dNLLB-200	22.59	8.85	29.51	18.82
lang2vec	21.85	<b>8.51</b>	33.38	19.27
NLLB	21.75	8.45	33.31	19.14
SuE	<b>21.90</b>	8.37	33.47	19.56
Random (init) (1 : 3)	21.88	8.47	<b>33.97</b>	<b>20.70</b>
Difference	-0.69	-0.34	+4.46	+1.88

Table 8: Mean BLEU scores with our methodology, as applied to Yoruba and Hungarian. **These scores represent preliminary results, where significantly less data (800 segments) were used to do CA-FT, and thus not directly comparable to Table 2.**

Error	👤	🌱	$s + t$	Notes
<b>Grammar Issues</b>				
<i>yo rann yo</i>	✗	✗	✗	Aspectual issue: English translation is passive, Creole is a voluntary activity
<i>te eksplòre</i>	✗	✗	✗	Aspectual error: English passive which also affects the sentence meaning
"many beautiful buildings to look at"	✓	✗	✗	🌱 and $s + t$ ungrammatical
"improve a considerable amount"	✓	✗	✗	🌱 and $s + t$ ungrammatical
"often provide some"	✓	✗	✓	🌱 ungrammatical
<i>fatìge mwens posib</i>	✗	✗	✗	👤 omits it; 🌱 and $s + t$ to "fatigued less possibly"
<i>konesans yo bay la</i>	✗	✗	✗	to "they give/provide" with no clear subject (passive would be better)
<i>mannyè efikas</i>	✗	✗	✗	missing preposition "in" ("in efficient manner")
<b>Translation Issues</b>				
Lexical errors				
<i>nimewotation</i>	✗	✗	✗	$s + t$ wrongly translates it to "naming system"
<i>benevola</i>	✗	✗	✗	"travel benevola" or "benevola travel" (in fact "charity")
<i>charabya</i>	✗	✗	✗	Means "nonsense", "gibberish"
<i>linets</i>	✗	✗	✗	👤 to "lenses"; 🌱 to "links"; $s + t$ not translating
<i>dyèl</i>	✗	✓	✗	👤 to "deals"; 🌱 correctly; $s + t$ to "dyed"
<i>patikilarite</i>	✓	✓	✗	$s + t$ to "patikilarity"
<i>kotidyen</i>	✓	✓	✗	$s + t$ not translated
<i>sètalò</i>	✗	✗	✗	👤 to "sevental"; 🌱 to "most"; $s + t$ to "satellite" (in fact "sometimes")
<i>bildinn</i>	✗	✗	✗	👤 to "image"; 🌱 to "figurine"; $s + t$ to "picture" (in fact "building")
<i>batiman</i>	✗	✗	✗	👤 and 🌱 to "vessels"; $s + t$ to "baths" (in fact "building")
<i>mannyè</i>	✓	✗	✗	👤 correctly to "manner"; 🌱 and $s + t$ to "manny"
<i>touch</i>	✗	✗	✗	translated to "touch" though it means "(piano) key"
<i>foumi militè</i>	✗	✗	✗	👤 and 🌱 to "military foam"; $s + t$ to "military fountains" (in fact "army ants")
<i>pyèj</i>	✗	✗	✗	👤 to "hostage"; 🌱 and $s + t$ to "pigeon" (in fact "trap")
<i>jako</i>	✗	✗	✗	👤 and 🌱 to "jaco"; $s + t$ omitted/misinterpreted (in fact "parrot")
<i>eskrimè</i>	✗	✗	✗	👤 and $s + t$ to "skirmish"; 🌱 to "skater" (in fact "fencer")
<i>parese</i>	✗	✗	✗	to "parasites" (in fact "sloths")
<i>pwomennen</i>	✗	✗	✗	to "promoted" (in fact "walking")
<i>eta espri</i>	✓	✓	✗	👤 and 🌱 correctly to "state of mind"; $s + t$ to "spirit state"
<i>Premye/Twazyèm Group</i>	✗	✗	✗	to "First/Third Group" (in fact "First/Third Class")
<i>mache</i>	✗	✗	✗	to "walk" (in fact "migrate" in context)
Other errors				
<i>kote pasaj enpótan yo</i>	✗	✗	✗	👤 and 🌱 too literal; $s + t$ also adding "their" instead of "the"

Table 9: Condensed categorization errors under various translation conditions. 👤 refers to Untrained Souping, 🌱 indicates IE Transfer, and  $s + t$  means  $s$  and  $t$  Adapters.