Automatic Speech Recognition and Query By Example for Creole Languages Documentation

Cécile Macaire, Didier Schwab, Benjamin Lecouteux, Emmanuel Schang

Automatic Speech Recognition and Query By Example for Creole Languages Documentation

Cécile Macaire[1], Didier Schwab[1], Benjamin Lecouteux[1], Emmanuel Schang[2]
[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
[2]LLL, UMR 7270, Univ. Orléans & CNRS
[1]first.last@univ-grenoble-alpes.fr
[2]emmanuel.schang@univ-orleans.fr

Abstract

We investigate the exploitation of self-supervised models for two Creole languages with few resources: Gwadloupéyen and Morisien. Automatic language processing tools are almost non-existent for these two languages. We propose to use about one hour of annotated data to design an automatic speech recognition system for each language. We evaluate how much data is needed to obtain a query-by-example system that is usable by linguists. Moreover, our experiments show that multilingual self-supervised models are not necessarily the most efficient for Creole languages.

1 Introduction

There is a long tradition of description of creole languages since, at least, the pioneering work of Hugo Schuchardt (1842-1927). Creole languages have sometimes been assigned a special role in linguistics: as a type of 'mixed languages', they are often considered as illustrating a break in language transmission and do not fit the generally assumed historical/genetic tree model[1]. Meanwhile they remain, for many of them, under-resourced languages. Gwadloupéyen, spoken mainly on Guadeloupe Island (France) by around 700.000 speakers, is a vigorous but largely under-equipped and under-resourced language. Morisien, (Mauritius Island) is spoken by approximately a million speakers. These two languages still suffer from a low social status and remain mainly spoken languages (rather than written).

The CREAM project aims at providing linguists with new methods for computational language documentation: Automatic Speech Recognition and keyword-spotting (Query-by-Example, Ram et al., 2020). The CREAM project teams up field linguists with computer scientists in order to address this low resource challenge.

Our contribution is twofold: 1) we introduce new methods for creole language documentation based on a combination of automatic speech recognition and keyword-spotting (in particular Query-by-example QbE) ; 2) our experiments shed new light on some key assumptions about creole languages, i.e. its distance from the lexifier language[2]. Gwadloupéyen (**gcf**) and Morisien (**mfe**) are two French-based creole languages. This category groups languages that share French as the most important part of their lexicon, but have a significantly different grammar.

We illustrate this phenomenon in (1-a.) and (1-b.) for gcf and mfe respectively, where most of the words (if not all) are clearly identifiable by a French speaker despite the difference in orthography:

(1) a. fo ou desann Gwadloup
need 2SG go.down Guadeloupe
'you've got to come to Guadeloupe"

b. Zan kontign reste.
John continue stay
'John continues to stay.' (from Henri and Kihm, 2015)

While these languages are well studied and vigorous, they are mostly spoken languages used in context of a dominant language: French for Gwadloupéyen (see Hazaël-Massieux, 1978; Managan, 2004, a.o.), French and English for Morisien (see Boswell, 2006; Rajah-Carrim, 2005, a.o.). Since they are spoken in two distinct linguistic and geographic areas (Lesser Antillean Island of Guadeloupe for Gwadloupéyen and Mauritius, in the Indian Ocean, for Morisien), there is no contact between these two languages, which makes them an interesting case study for a comparison (no con-

[1]But see DeGraff (2004) or Corcoran (2001) for a severe criticism of the "creation myth".

[2]The language which provides the most part of the lexicon.

tact[3], lexicon based on French, different grammars).

2 Transcribing Creole Languages

Since Gwadloupéyen and Morisien are mostly spoken languages, their written form is not stable. There are resources such as dictionaries and grammars for both languages (Tourneux and Barbotin, 2009; Ludwig et al., 1990; Damoiseau, 2012; Police-Michel et al., 2012; Baker, 1972; Baker and Hookoomsing, 1987), but writing in creole and transcribing spoken speech are two separate tasks.

In the context of diglossia, code-switching is very frequent (see Auckle, 2015; Jeannot and Jno-Baptiste, 2008; Hazaël-Massieux, 1978) and obviously causes problems for an automatic transcription task.

We focused here mainly on Gwadloupéyen and we identified three main problems with the transcriptions available in (Glaude, 2013).

First, several words are transcribed in two different forms: *anko* vs *ankò* 'again', *apré* vs *après* 'after', *bitin* vs *biten* 'thing'.

Second, the transcriber hesitates between a transcription in French or in Creole:

(2) modes de cuisson qui adaptés osi
fr fr fr fr fr cr
methods of cooking that adapt too
'cooking methods which are adapted too'

As shown in (2), the transcriber chose in this segment to write a large segment in French (fr), except for the word *osi* 'too, also', which is pronounced the same way in French and Gwadloupéyen (i.e. [osi]) but written *aussi* in French. However, one can wonder why *adapté* is not written in creole (no number agreement then), why *qui* is not written *ki* and, perhaps *modes de cuisson* transcribed *mode dé kwison* (since *é* in creole can be pronounced [ø]).

And last, the transcriber chose to transcribe in the proper creole form (identified as basilectal) while the speaker pronounced a word quite similar to its form in French: transcribed *dantis* but pronounced as in French *dentiste*.

Creole languages are known to have a large range of variation, often described as the 'Creole continuum', (see Bickerton, 1973; Mufwene, 1997; Winford, 1997, among many). This fact has even been theorized as a historical evolution towards the lexifier language, but see Mufwene (1997); Aceto (1999); Prudent (1999); Aboh (2015) for a more nuanced approach or a radical critic of this approach (DeGraff, 2004). In any case, Creole variations is a source of difficulty for ASR systems.

In order to efficiently correct these errors and to allow the linguist to search for a word (i.e. a segment of speech) in the corpus independently of its transcription, we designed an experiment of keyword spotting (**QbE**). This task is in line with Bird (2021), and is brought into action when there is a need for the linguist to verify or correct the transcription.

Speech processing for creole languages has not received much attention so far. For Gwadloupéyen, Delumeau (2006) is, to our knowledge, the only relevant work in NLP, but it does not address speech recognition. For Haitian Creole, Breiter (2013) explores speech recognition but Haitian and Gwadloupéyen are clearly distinct languages. For Morisien (Noormamode et al., 2019) is a recent initiative for creating a Creole speech engine. However, it does not seem to address the same tasks as this work.

3 ASR with Self-supervised Learning

Self-supervised learning (SSL) is the task of learning powerful representations from huge unlabeled data (called pretraining) to recognize and understand patterns from a less common problem (called fine-tuning). Recent work focused on speech data have reported impressive results for representation learning, and more specifically improved performance on downstream tasks for ASR in low-resource contexts (Baevski et al., 2019; Kawakami et al., 2020). These work are based on the Wav2Vec2.0 (Baevski et al., 2020) model.

In our approach, we consider 2 models : 1) *XLSR-53* (Conneau et al., 2021), a multilingual pretraining of Wav2Vec2.0 model on 53 languages with more than 56k hours of unlabeled speech data (*XLSR-53*) which has been shown to construct better speech representations for cross-lingual transfer (Conneau et al., 2021); 2) *LeBenchmark* (Evain et al., 2021a,b), a French-based Wav2Vec2.0 model with the assumption that these creoles are closely related to French.

4 Query by Example

Query by Example (QbE) consists in detecting specific words in speech recordings thanks to the use of speech recognition approaches. Keywords are

[3] And no mutual understanding (Chaudenson, 2004).

defined according to the user's request. Within the scope of this work, our keyword spotting approach firstly uses self-supervised learning models to predict the word in a speech segment. In the second phase, it searches for the prediction in a set of transcriptions.

5 Methodology

Dataset We consider two creole languages: Gwadloupéyen (gcf, 80 min and 5 speakers) and Morisien (mfe, 60 min and 2 speakers). Corpora are provided by Glaude (2013) for gcf and by courtesy of Dr. Tonjes Veenstra[4] for Morisien. Both corpora contain paired data of spontaneous speech with corresponding transcriptions.

Pre-processing for Fine-tuning Each audio recording is segmented into small segments, each corresponding to a sentence. Audio segments are mono, with a sampling frequency of 16 kHz. The pre-processing of textual data involves the deletion of punctuation marks, and a harmonization of specific characters (lowercase, the substitution of accentuated vowels such as 'à' into 'a', 'ê' into 'e', ...). For each experiment, we split the data into train, validation and test sets with a ratio of 80/8/12. Details about the datasets are given in Appendice A.1.1.

Implementation Details The fine-tuning is performed using the Wav2Vec2.0 model (Wolf et al., 2020). We used two pretrained models available in HuggingFace (Wolf et al., 2020): *XLSR-53-large* multilingual model (Conneau et al., 2021), and *LeBenchmark/wav2vec2-FR-7K-large* (Evain et al., 2021b). Hyperparameters are the same as Conneau et al. (2021), except for the batch size, set to 8 due to memory limitations (see Appendice A.1.2). For LM rescoring, we build 3-gram language models (LM) using KenLM (Heafield, 2011) on the training transcriptions (see Table 1 for details). Results are generated with a CTC beam search decoder (Graves et al., 2006).

Query by example We create a set of speech segments for Gwadloupéyen language, with each corresponding to a word. The utterances are carefully chosen: we extracted pieces of signals (outside the train and test sets) that could be found in the test data to simulate the work of a linguist. We used the fine-tuned models to generate the corresponding transcription of an audio segment. In this part, we do not decode with a LM to get closer to the signal. We base our approach on the Smith-Waterman algorithm (Smith and Waterman, 1981). This method provides an optimal local alignment between two given sequences by looking at matching areas (Lecouteux et al., 2012). QbE approach was performed with non-optimized weights by default (substitution, insertion, and deletion set to 1).

Language	# 1gram	# 2grams	# 3grams	Perplexity (%)
gcf	1584	6530	9431	112.82
mfe	1293	5274	7308	141.14

Table 1: Statistics and perplexity of 3-gram LM of Gwadloupéyen (gcf) and Morisien (mfe) languages.

6 Results

Automatic Transcription Performance We evaluate the fine-tuned models performance using the Word Error Rate (WER) and the Character Error Rate (CER) with and without a 3-gram LM. Results are displayed in Table 2.

For both creole languages, models using the *LeBenchmark* model perform better in comparison to the multilingual model with a gain of over 5 to 8 percentage points (35.96%/40.68% WER for gcf, 36.19%/44.66% WER for mfe). To support our results, we performed cross-validation on the Gwadloupéyen corpus (see Appendice A.2.1). We conducted complementary experiments to assess the model's performance on data from an unseen speaker (see Appendice A.2.2).

Query by example and ASR In an attempt to know how much data is needed to get satisfactory performance (usability in the context of linguistic fieldwork), and whether the approach can be generalized to other related creole languages, we conducted several fine-tuning runs with different training dataset sizes (from 10 min to 70 min), only on Gwadloupéyen data[5]. The WER on the test data is given for each fine-tuned model in Figure 1. We observe impressive results with less

[4]The data on Kreol Morisien were collected by Tonjes Veenstra within the context of the A02-project, entitled "Speaker's choices in a creole context: Bislama and Morisien", of the CRC 1412 on Register, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412, 416591334. The data will be made publicly available at the end of the project.

[5]Audio segments were selected by an expert of Gwadloupéyen.

Model	Training size (in min)	Pretrained model	LM	dev		test	
				WER (%)	CER (%)	WER (%)	CER (%)
gcf_xlsr	68	facebook/wav2vec2-large-xlsr-53	-	47.58	22.60	40.68	17.81
			3-gram	-	-	37.91	18.59
gcf	68	LeBenchmark/wav2vec2-FR-7K-large	-	39.50	17.89	35.96	15.86
			3-gram	-	-	**34.74**	**16.96**
mfe_xlsr	52	facebook/wav2vec2-large-xlsr-53	-	48.08	21.56	44.66	20.06
			3-gram	-	-	41.60	20.12
mfe	52	LeBenchmark/wav2vec2-FR-7K-large	-	41.44	18.23	**36.19**	**16.70**
			3-gram	-	-	38.83	18.03

Table 2: Word Error Rate (WER) and Character Error Rate (CER) on different creole languages when fine-tuning the Wav2Vec2.0 model with multilingual (*XLSR-53*) and monolingual (*LeBenchmark/wav2vec2-FR-7K-large*) models. The WER and the CER are given with and without a 3-gram LM on the test sets.

than 1 hour of paired audio and transcriptions. Our query by example approach, over a set of 13 Gwadloupéyen audio segments, gives precision and recall scores of over 70% (84.52%/84.94% with the Gwadloupéyen model trained on 60 minutes). In addition, using the model trained with only 10 minutes of data gives very good performance (83.33% Precision/74.36% Recall), which shows its effectiveness in low resource contexts.

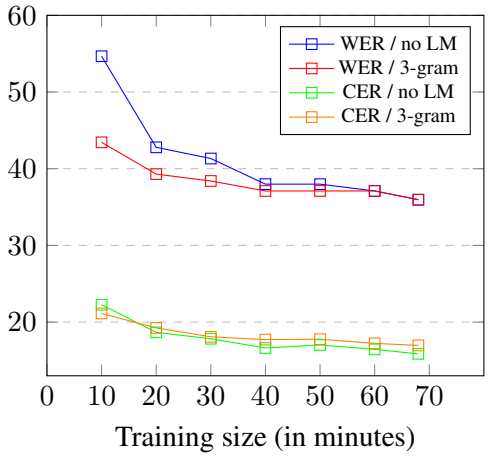

Figure 1: WER and CER (%) with respect to different training sizes (in minutes) when fine-tuning LeBenchmark pretrained model on the Gwadloupéyen corpus. The WER and the CER are given on the test sets with (in red and orange) or without a 3-gram LM (in blue and green).

7 Discussion

Field linguists from the CREAM project evaluate very positively the results given in Table 2. As shown in Appendice A.3, the automatic transcription can already save a huge amount of time and is accurate enough to allow for a fast manual correction.

Fine-tuned model	Precision (%)	Recall (%)	F-measure (%)
gcf_10	83.33	74.36	78.59
gcf_20	83.33	76.28	79.65
gcf_30	72.50	79.49	75.83
gcf_40	75.00	74.36	74.68
gcf_50	66.67	66.67	66.67
gcf_60	**84.52**	**84.94**	**84.73**

Table 3: Precision, Recall and F-measure computed on the Qbe results of 13 Gwadloupéyen audio segments when using the fine-tuned *gcf* models trained with 10 (gcf_10) to 60 minutes (gcf_60) of training data to predict the utterance. Audio segments contain single words (e.g. 'dépi', 'fè') and multiple words (e.g. 'an pa sav', 'nou ka rivé').

Moreover Table 2 sheds new light on the question of the link between a Creole language and the so-called 'lexifier' language (French for Gwadloupéyen and Morisien). It has been hypothesized that creole languages form a special typological class of languages (see Bakker et al., 2017, for a detailed discussion) or even a class of simple languages (see McWhorter, 2001). At the phonological level, creole languages are supposed to have phonological inventories that are distinct from those of their lexifiers. However, our results show that a model pretrained on French performs better than a model trained on a typologically wide sample (53 languages are taken into account in *XLSR-53*, including Haitian, which is a French-based creole language). If creole languages were so different from their lexifier languages (French in our case), we should expect a better performance on a 53 languages pretrained model. Interestingly, for Gwadloupéyen and Morisien, French is obviously the common connection. But in the case of

Morisien, most speakers are also fluent in English (Atchia-Emmerich, 2005), which could also have had an impact on the results. As underlined in Atchia-Emmerich (2005), French still remains an important language for Mauritians, and English, despite its high social prestige, does not have a significant impact on Morisien.

8 Conclusion and perspectives

Of course, an ASR system cannot solve the problems that the human transcribers have not solved, i.e. the choice of transcribing a word in French or in Creole (code-switching or not)[6].

Our results show that QbE can complement ASR and provide an easy way to scan the corpus for relevant examples. We found that a model pretrained on French performed better for Gwadloupéyen and Morisien than a model pretrained on a large typological set of languages[7].

For future work, we intend to apply the same method on English-based creole languages (such as Jamaican Creole) and Portuguese-base creoles (Kriol of Guinea-Bissau), to allow for a comparison and a generalization.

Acknowledgments

The CREAM project (Documentation des Langues CREoles Assistée par la Machine) is funded by the ANR (Agence Nationale de la Recherche, CS-38, 2020-2024). We would like to thank Dr. T. Veenstra (ZAS, Berlin) and Dr. F. Henri (University at Buffalo) for sharing with us their data on Morisien.

References

Enoch Oladé Aboh. 2015. *The emergence of hybrid grammars: Language contact and change*. Cambridge University Press.

Michael Aceto. 1999. Looking beyond decreolization as an explanatory model of language change in creole-speaking communities. *Journal of Pidgin and Creole Languages*, 14(1):93–119.

Bilkiss Atchia-Emmerich. 2005. La situation linguistique à l île maurice. les développements récents à la lumière d une enquête empirique. *Université de Nuremberg, dissertation inaugurale à la Faculté de philosophie II (science de la langue et de la littérature, 20 janvier 2005*.

Tejshree Auckle. 2015. *Code switching, language mixing and fused lects: language alternation phenomena in multilingual Mauritius*. Ph.D. thesis, University of South Africa.

Alexei Baevski, Michael Auli, and Abdelrahman Mohamed. 2019. Effectiveness of self-supervised pretraining for speech recognition. *arXiv preprint arXiv:1911.03912*.

Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. 2020. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, 33:12449–12460.

Philip Baker. 1972. *Kreol: a description of Mauritian Creole*. C. Hurst.

Philip Baker and Vinesh Y Hookoomsing. 1987. *Diksyoner kreol morisyen: Dictionary of Mauritian Creole*. Editions L'Harmattan.

Peter Bakker, Finn Borchsenius, Carsten Levisen, and Eeva Sippola. 2017. *Creole studies–phylogenetic approaches*. John Benjamins Publishing Company.

Derek Bickerton. 1973. The nature of a creole continuum. *Language*, pages 640–669.

Steven Bird. 2021. Sparse transcription. *Computational Linguistics*, 46(4):713–744.

Rosabelle Boswell. 2006. *Le malaise creole: Ethnic identity in Mauritius*, volume 26. Berghahn Books.

Wojtek Breiter. 2013. Rapid bootstrapping of haitian creole large vocabulary continuous speech recognition. Master's thesis, Karlsruher Institut fur Technologie.

Robert Chaudenson. 2004. *La créolisation: théorie, applications, implications*. Editions L'Harmattan.

Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, and Michael Auli. 2021. Unsupervised Cross-Lingual Representation Learning for Speech Recognition. In *Proc. Interspeech 2021*, pages 2426–2430.

Chris Corcoran. 2001. Creoles and the creation myth: A report on some problems with thelinguistic use of 'creole'. In *Papers from the 36th Meeting of the Chicago LinguisticSociety. Chicago: Chicago Linguistics Society*.

Robert Damoiseau. 2012. Syntaxe créole comparée. *Karthala et CNDP-CRDP edition*.

Michel DeGraff. 2004. Against creole exceptionalism (redux). *Language*, 80(4):834–839.

[6] See (Delumeau, 2006) for a synthesis of the discussion about writing and orthography in gcf, and (Vaillant and Léglise, 2014) for the difficulties of language identification and code-switching in the context of creole languages.

[7] Tools are provided in https://github.com/macairececile/ASR-QbE-creole

Fabrice Delumeau. 2006. *Une description linguistique du Creole Guadeloupéen dans la perspective de la genération automatique d'enonces.* Ph.D. thesis, Université de Nanterre-Paris X.

Solène Evain, Ha Nguyen, Hang Le, Marcely Zanon Boito, Salima Mdhaffar, Sina Alisamir, Ziyi Tong, Natalia Tomashenko, Marco Dinarelli, Titouan Parcollet, Alexandre Allauzen, Yannick Estève, Benjamin Lecouteux, François Portet, Solange Rossato, Fabien Ringeval, Didier Schwab, and laurent besacier. 2021a. Task Agnostic and Task Specific Self-Supervised Learning from Speech with LeBenchmark. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Solène Evain, Ha Nguyen, Hang Le, Marcely Zanon Boito, Salima Mdhaffar, Sina Alisamir, Ziyi Tong, Natalia Tomashenko, Marco Dinarelli, Titouan Parcollet, Alexandre Allauzen, Yannick Estève, Benjamin Lecouteux, François Portet, Solange Rossato, Fabien Ringeval, Didier Schwab, and Laurent Besacier. 2021b. LeBenchmark: A Reproducible Framework for Assessing Self-Supervised Representation Learning from Speech. In *Proc. Interspeech 2021*, pages 1439–1443.

Herby Glaude. 2013. Corpus créoloral. oai:crdo.vjf.cnrs.fr:crdo-GCF. *SFL Université Paris 8 - LLL Université Orléans.*

Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning*, pages 369–376.

Guy Hazaël-Massieux. 1978. Approche sociolinguistique de la situation de diglossie français-créole en guadeloupe. *Langue française*, (37):106–118.

Kenneth Heafield. 2011. Kenlm: Faster and smaller language model queries. In *Proceedings of the sixth workshop on statistical machine translation*, pages 187–197.

Fabiola Henri and Alain Kihm. 2015. The morphology of tam marking in creole languages: a comparative study. *Word Structure*, 8(2):248–282.

Béatrice Jeannot and Durizot Jno-Baptiste. 2008. L'enseignement du français en contexte diglossique guadeloupéen: état des lieux et propositions. *Former les enseignants du XXIème siècle dans toute la francophonie*, page 61.

Kazuya Kawakami, Luyu Wang, Chris Dyer, Phil Blunsom, and Aaron van den Oord. 2020. Learning robust and multilingual speech representations. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1182–1192, Online. Association for Computational Linguistics.

Benjamin Lecouteux, Georges Linares, and Stanislas Oger. 2012. Integrating imperfect transcripts into speech recognition systems for building high-quality corpora. *Computer Speech and Language*, 26(2):67 – 89. (Impact-F 1.46 estim. in 2012).

Ralph Ludwig, Danièle Montbrand, Hector Poullet, and Sylviane Telchid. 1990. Abrégé de grammaire du créole guadeloupéen. *Dictionnaire créole français (Guadeloupe), avec un abrégé de grammaire créole et un lexique françaiscréole*, pages 17–38.

Jane Kathryn Managan. 2004. *Language choice, linguistic ideologies and social identity in Guadeloupe*. New York University.

John H McWhorter. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology*, 5.

Salikoko S. Mufwene. 1997. Introduction: Understanding speech continua. *World Englishes*, 16(2):181–184.

Wajiha Noormamode, Baby Gobin-Rahimbux, and Mohammad Peerboccus. 2019. A speech engine for mauritian creole. In *Information Systems Design and Intelligent Applications*, pages 389–398. Springer.

Daniella Police-Michel, Arnaud Carpooran, and Guilhem Florigny. 2012. *Gramer kreol morisien: volim I. Dokiman referans*. Akademi Kreol Morisien, Ministry of Education and Human Resources.

Hector Poullet, Sylviane Telchid, and Daniele Montbrand. 1984. *Dictionnaire des expressions du créole guadeloupéen*. Hatier Antilles.

Lambert-Félix Prudent. 1999. *Des baragouins à la langue antillaise: Analyse historique et sociolinguistique du discours sur le créole*, volume 1. Caribeennes.

Aaliya Rajah-Carrim. 2005. Language use and attitudes in mauritius on the basis of the 2000 population census. *Journal of multilingual and multicultural development*, 26(4):317–332.

Dhananjay Ram, Lesly Miculicich, and Hervé Bourlard. 2020. Neural network based end-to-end query by example spoken term detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:1416–1427.

Temple F Smith and Michael S Waterman. 1981. Identification of common molecular subsequences. *Journal of molecular biology*, 147(1):195–197.

Sylviane Telchid, Hector Poullet, and Frédéric Anciaux. 2009. *Le Déterville: dictionnaire français-créole*. PLB Editions.

Henry Tourneux and Maurice Barbotin. 2009. *Dictionnaire pratique du créole de Guadeloupe (Marie-Galante)*. KARTHALA Editions.

Pascal Vaillant and Isabelle Léglise. 2014. À la croisée des langues. annotation et fouille de corpus plurilingues. *Revue des Nouvelles Technologies de l'Information*, pages 81–100.

Donald Winford. 1997. Re-examining caribbean english creole continua. *World Englishes*, 16(2):233–279.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

A Appendices

A.1 Appendice A: Implementation details

A.1.1 Datasets

Corpus	Train	Dev	Test	OOV (%)
gcf	68	4	8	24.68
mfe	52	3	5	24.82

Table 4: Train, dev and test sizes in minutes of the Creole corpora used to fine-tune Wav2Vec2.0 pretrained models, as well as the percentage of out-of-vocabulary words.

A.1.2 Hyperparameters

Hyperparameters are given in Table 5.

A.2 Appendice B: Complementary results

A.2.1 Cross-validation results

Results of the cross-validation experiments are printed in Table 6.

A.2.2 Experiments on an unseen speaker in the training set

In the experiments on Gwadloupéyen, the train/test split has been randomly performed, speakers can be both in the train and the test sets. This is frequent in speech recognition evaluations. It corresponds to our use case (in a low-resource context and working with few recordings). We conduct a complementary experiment that excludes a speaker from the training set and evaluates the performance of the fine-tuned model on two test sets: with or without speaker audio segments. Results are displayed in Table 7.

The results show that the model has a close word and character error rates on the audio segments of a speaker not seen in the training data (40.66% WER against 36.46% WER). When decoding with a 3-gram language model, the WER on the test data of the unseen speaker is degraded by one percentage point compared to the other test set (37.62%/36.29% WER).

A.3 Appendice C: Sample of Error Analyses for Gwadloupéyen

False negatives In some cases the manual transcription (Ref) was incorrect and the model (Hyp) provides an accurate hypothesis. Among others, these are several problems:

Missing word 'la' in the manual transcription:

```
Ref: sé timoun pa ni pon rèspè
Hyp: sé timoun la pa ni pon
     respè
```

Dysfluences 'é' (hesitation) are missing in Ref:

```
Ref: donk chak ritm la
ka espliké on biten
Hyp: donk é chak ritm la
ka espliké on biten
```

The Ref version makes an **inappropriate elipsis** (grammatical but not in the recording):

```
Ref: <pou pé> négosyé
sé péyi [...]
Hyp: <pou ou pé> négosi
sé péyi la [...]
```

The Hyp version detects the correct spelling of the pronoun (atone vs tonic):

```
Ref: <mwen> pa ka di lafrans
     [...]
Hyp: <an> pa ka di la frans [...]
```

Non decidable Some words are not present in (Poullet et al., 1984; Telchid et al., 2009; Tourneux and Barbotin, 2009) and the Hyp is rather correct:

```
Ref: tandis ké gwada
Hyp: tandiské gwada
```

Since 'tandis' cannot occur without 'ké', the Hyp makes a correct guess.

In the cases where it is impossible to decide if the segment is in French (with a creole accent) or in Creole, the Hyp is not faulty:

parameter	value
pretrained model	wav2vec2-large-xlsr-53 LeBenchmark/wav2vec2-FR-7K-large
attention_dropout	0.1
hidden_dropout	0.1
feat_proj_dropout	0.1
mask_time_prob	0.075
layerdrop	0.1
ctc_loss_reduction	mean
train_batch_size	8
num_train_epochs	60
fp16	True
learning_rate	3e-4

Table 5: Value of the hyperparameters used to fine-tune the Wav2Vec2.0 model on Gwadloupéyen and Morisien datasets.

Model	**WER (%)**		**CER (%)**	
	None	3-gram	None	3-gram
split 1	34.64	32.59	16.62	15.57
split 2	34.65	33.60	14.83	15.24
split 3	34.85	33.92	13.83	15.07
split 4	35.18	35.61	14.36	16.15
split 5	35.48	34.74	15.33	16.38
split 6	36.36	37.48	16.17	18.35
split 7	35.50	36.10	15.21	16.23
split 8	37.55	37.55	16.76	18.00
split 9	35.37	36.25	16.07	17.06

Table 6: Cross validation on Gwadloupéyen dataset when fine-tuning Wav2Vec2 model with the *LeBenchmark/wav2vec2-FR-7K-large* pretrained model. 9 different datasets were created from the Gwadloupéyen dataset, with 68 minutes of training data, 4 minutes of validation data and 8 minutes of test data. The WER and the CER are given with and without a 3-gram language model on the test sets.

```
Ref: é le grand bourg exactement
Hyp: é le gran bou egzaktéman
```

Grand Bourg is the French name for a town of Marie-Galante Island, and the adverb 'exactement/egzaktéman' can be pronounced in the same way in fr and gcf.

```
Ref: é on avansé o nivo <mantal>
     paské
Hyp: on avansé o nivo <mental>
     paské
```

Here, 'mental' (fr) and 'mantal' (gcf) have the same spelling.

```
Ref: a sé jèn la èvè <lentènèt> é
     tou sa
Hyp: sé jann la èvè <lintenèt> é
     tou sa
Ref: an plas an fòs mèm an plas
<sitou> an frans
Hyp: an plas an fòs mèm an plas
<soutou> an frans
```

Both forms can be found.

A.4 Appendice D: Query-by-Example outputs

Correct QbE The *gcf_60* fine-tuned model predicts the word 'depi' for a given speech segment. The QbE approach extracts several results where this keyword is seen in a transcription, one of which is the following:

```
Query: 13 depi 18
          ||||
Ref  :  1 depi 6

Score: 12
Matches: 6 (100.0%)
Mismatches: 0

File name: 1016_273.wav
Complete sentence:
sa vle di ke depi le le an rantre
tale a kaz prepare sak an mwen
```

Incorrect QbE The *gcf_60* fine-tuned model predicts the word 'pasew' for a given speech segment. In this case, the prediction is incorrect ('pase' is the keyword we are looking for).

```
Query: 31 paske-  37
         |||| | |
Ref  : 1  pas-ew  7
```

Model	Training size (in min)	LM	dev		test no speaker		test speaker	
			WER (%)	CER (%)	WER (%)	CER (%)	WER (%)	CER (%)
gcf_speaker	64	-	42.48	19.02	36.46	16.37	40.66	17.12
		3-gram	-	-	36.29	17.55	37.62	19.05

Table 7: Word Error Rate (WER) and Character Error Rate (CER) on gwadeloupean language when fine-tuning the Wav2Vec2.0 model with *LeBenchmark/wav2vec2-FR-7K-large* model by excluding one speaker from the train set. The WER and the CER are given with and without a 3-gram LM on two test sets: one with speaker audio segments (test speaker, 7.5 minutes) and one without (test no speaker, 5 minutes).

```
Score: 10
Matches: 6 (75.0%)
Mismatches: 2
Path of the file: 1041_0194.wav

Complete sentence: tou se moun la
ki ka tout moun paske tout moun
ka rankontre obstak
```