



# WolBanking77: Wolof Banking Speech Intent Classification Dataset

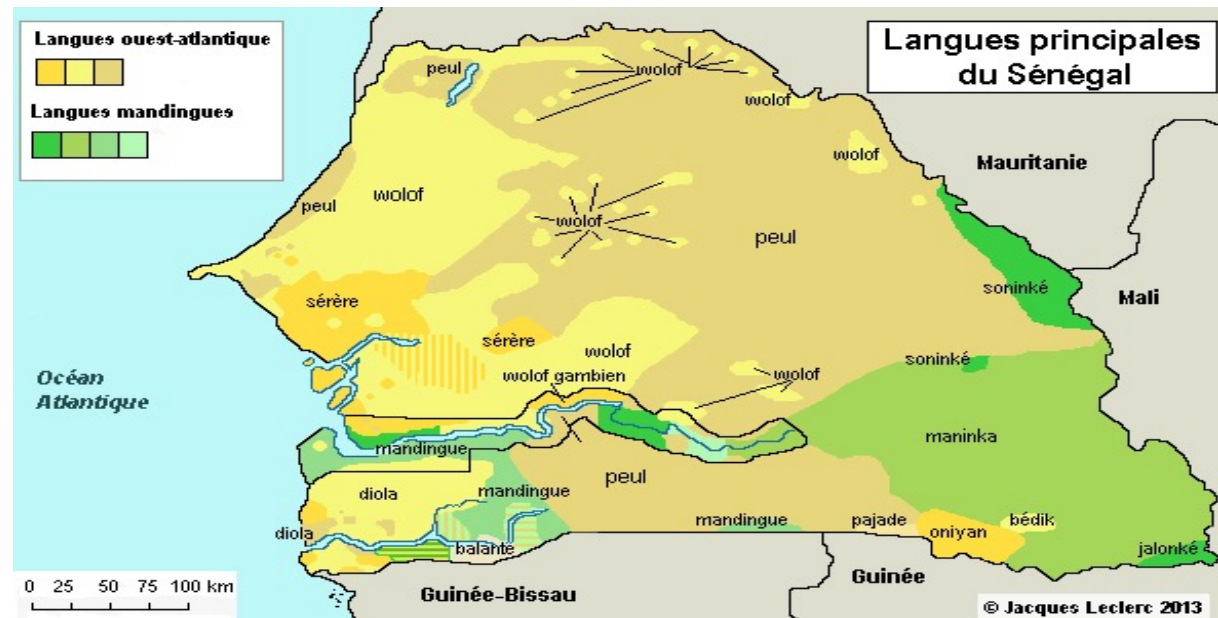
Abdou Karim Kandji<sup>\*1</sup>, Frédéric Precioso<sup>2</sup>, Cheikh Ba<sup>1</sup>, Samba Ndiaye<sup>3</sup>, Augustin Ndione<sup>3</sup>

<sup>1</sup>University of Gaston Berger (UGB), <sup>2</sup>Université Côte d'Azur (UniCA), <sup>3</sup>Cheikh Anta Diop University (UCAD)



## Wolof Language

- Wolof is spoken in West African region (Senegal, Gambia and Mauritania) by more than 10 million people.
- Most of the population in Senegal speaks Wolof (around 90%) which is essentially a language of communication between several ethnic groups.



Source : Jacques Leclerc (2023). "Langues principales du Sénégal"

## Context

- Why ? The rate of the illiterate population is estimated at **48%** in **Senegal** according to the World Bank
- With **+15 million Internet subs.**, **97.72%** represent mobile users
- WhatsApp or Facebook Messenger enable the exchange of voice messages between users regardless of their level of education
- Voice processing applications to facilitate technology access & financial inclusion in Senegal

## Motivation

We present **WolBanking77**, an Intent Classification Dataset released to address the gap in intent classification for low-resource languages, particularly in Wolof language, with experiments conducted on text and voice models showing promising results.

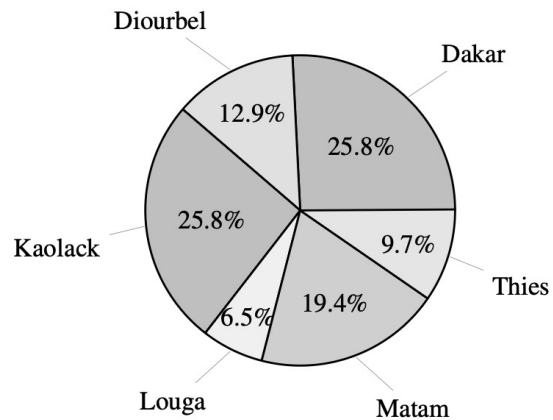
Dataset	Num Utt	Banking Domain
ALFFA	18,000	✗
Kallaama	56,693	✗
INJONGO	3,200	✓
<b>WolBanking77(Ours)</b>	9,791	✓

## Main contributions

- An audio dataset with **263** sentences (**4 hours**) covering **10** intents in the banking and transport domains
- A text dataset with **9,791** sentences translated from English to French and Wolof covering **77** intents in banking domain
- Experimental benchmarks on Automatic Speech Recognition and Intent Detection tasks
- Datasheets for WolBanking77

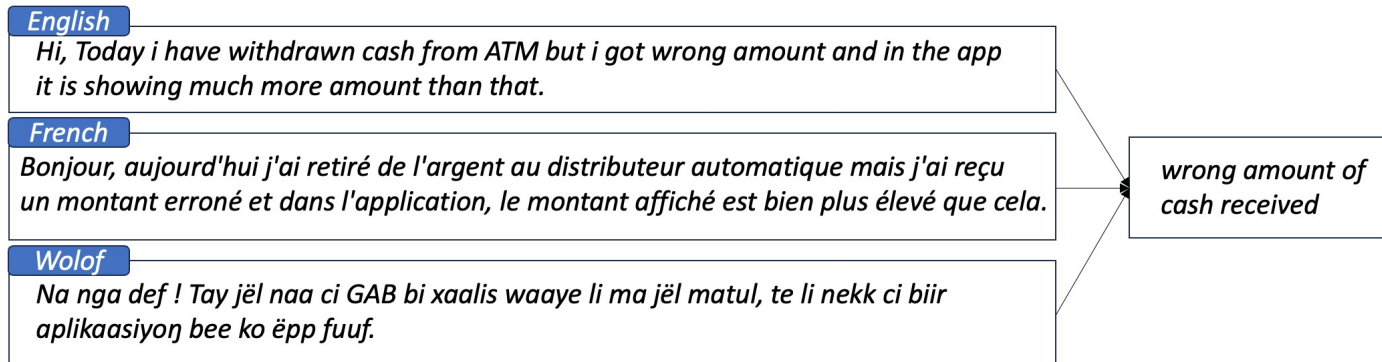
## WolBanking77

SPLIT	5k sample	all
TRAIN SET	4,000	7,832
TEST SET	1,000	1,959



intent	domain
BALANCE	e-banking
CASH_DEPOSIT	
FREEZE	
LATEST_TRANSACTIONS	
PAY_BILL	
OPEN_ACCOUNT	
TRANSFER_MONEY	transport
AMOUNT	
BUS_RESERVATION	
TECHNICAL_VISIT	

## Main Task: Customer Intent Detection

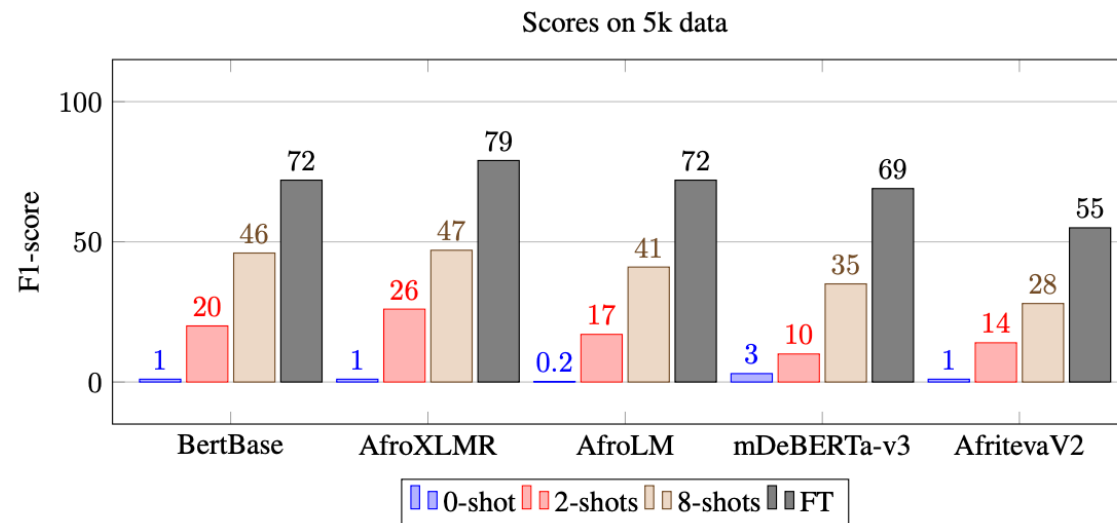


**WolBanking77** is also suitable for :

- Machine Translation in French & Wolof
- Automatic Speech Recognition in Wolof
- Comparing different machine learning models for Intent Classification.

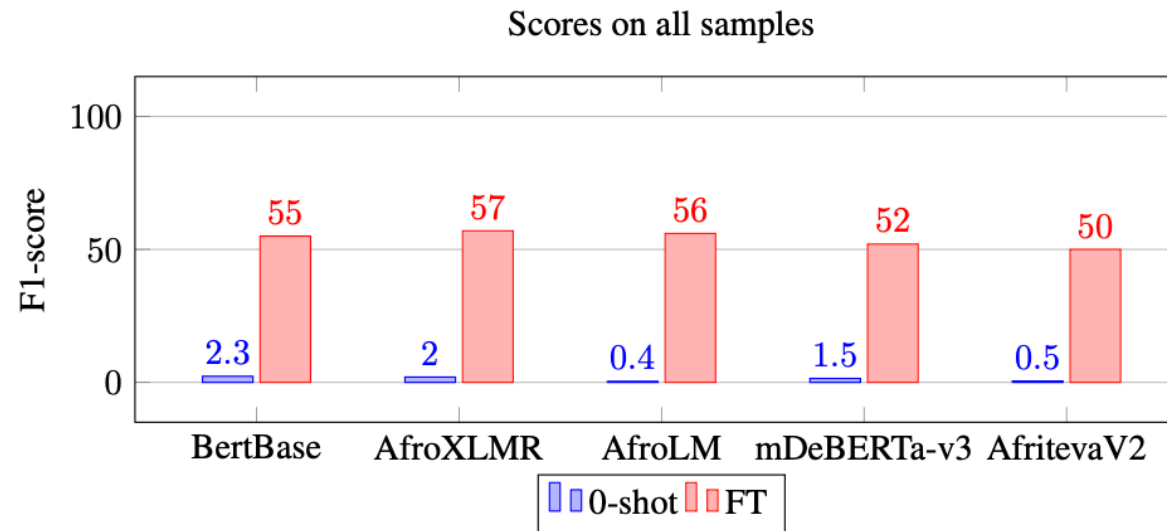
## Intent Detection results

Model	5k samples			All samples		
	Precision	Recall	F1	Precision	Recall	F1
Llama-3.2-1B-Instruct	0.74	0.71	0.72	0.52	0.45	0.46
Llama-3.2-3B-Instruct	0.76	0.75	0.75	0.56	0.55	0.55





## Intent Detection results



## ASR results

Model	Parameters	Training time	Steps	WER
Phi-4-multimodal-instruct	5.6B	32	1000	3.1%
Distil-whisper-large-v3.5	756M	44	1000	4.63%
<b>Canary-1b-flash</b>	1B	20	1000	<b>0.59%</b>

## Thank you

Download the dataset on GitHub:

<https://github.com/abdoukarim/wolbanking77>

