

# Collecting Resources in Sub-Saharan African Languages for Automatic Speech Recognition: a Case Study of Wolof



Elodie Gauthier<sup>1</sup>, Laurent Besacier<sup>1</sup>, Sylvie Voisin<sup>2</sup>, Michael Melese<sup>3</sup> and Uriel Pascal Elingui<sup>4</sup>

<sup>1</sup>Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France

<sup>2</sup>Laboratoire Dynamique Du Langage (DDL), CNRS - Université de Lyon, France

<sup>3</sup>Addis Ababa University, Ethiopia

<sup>4</sup>Voxygen SAS, Pleumeur-Bodou, France & Dakar, Sénégal

laurent.besacier@imag.fr



## The ALFFA Project

African Languages in the Field : speech Fundamentals and Automation

- African Languages and Information Technologies
- Address under-resourced languages with immediate or short-term potential (economical, strategic)
- Data collection methodology
- ASR and TTS technologies
- Micro-application prototypes for mobile phones
- Inter-disciplinary project : gather technology experts and fieldwork linguists/phoneticians
- Target languages : Hausa, Bambara, Wolof, Pulaar, Swahili, Amharic
- ASR systems for Swahili, Hausa and Amharic designed to date

Task	LM corpus (#words)	AM training	Evaluation set	WER (%)
Swahili broadcast news	28M	10h	1.8h	20.7
Hausa read speech	42k	7h	1h	10.0
Amharic read speech	4M	20h	2h	8.7 <sup>1</sup>

## Collecting data in Wolof (spoken in Senegal)

### Key figures

- 4 millions of native speakers
- 90 % of the Senegalese people speak and understand Wolof
- 40 % use it as their mother tongue

### Major issues

- very under-resourced language
- very few electronic documents available
- no real rules for the orthography since Wolof is not learnt at school

### Collecting text

- 148k words initially available
- 500k words crawled on the Web

Text	#utterances	#tokens
Universal Declaration of Human Rights	112	1,923
Silo's Message	602	10,443
The Bible	14,474	185,064
Wikipedia	10,738	311,995
<b>Total</b>	<b>25,926</b>	<b>509,425</b>

### Collecting audio

- 10 native Wolof speakers
- 18,000 utterances recorded
- 21h21 mins of signal

Set	Male	Female	#utterances	#tokens	Duration
Training	8	6	13,998	132,963	16 h 49 mins
Development	1	1	2,000	18,790	2 h 12 mins
Testing	1	1	2,000	18,843	2 h 20 mins
<b>Total</b>	<b>10</b>	<b>8</b>	<b>17,998</b>	<b>170,596</b>	<b>21 h 21 mins</b>

## First ASR system for Wolof

### Resources built

- 2 languages models (LMs) trained :
  - > 3-gram LM trained on 106k words (*called LM1*)
  - > 3-gram interpolated LM trained on 600k words (*called LM2*)
- pronunciation dictionary of 32,039 entries
- 3 acoustic modeling techniques built

Wolof ASR systems performance for different AMs and LMs – with speaker adaptation

Acoustic model	WER (%)			
	LM1 (~106k)		LM2 (~600k)	
	dev	test	dev	test
HMM/GMM	33.51	37.95	31.70	35.97
<i>no diacritic</i>	31.60	36.20	29.70	34.10
SGMM+MMI	30.37	35.24	28.56	33.56
<i>no diacritic</i>	28.40	33.50	26.60	31.70
DNN+sMBR	29.10	35.45	<b>27.21</b>	33.63
<i>no diacritic</i>	27.20	33.60	25.10	31.70

[https://github.com/besacier/ALFFA\\_PUBLIC](https://github.com/besacier/ALFFA_PUBLIC)

<sup>1</sup> Actually, it is a morph error rate.