

Un corpus du kriol à l'épreuve du TAL pour l'étude de la variation



Flora Badin & Jean-Louis Rougé

LLL UMR7270, Université d'Orléans, 10 rue de Tours, 45065 Orléans, France

LES CRÉOLES, LE KRIOL EN GUINÉE BISSAU ET EN CASAMANCE

Les créoles sont nés du contact entre différentes langues. L'une d'entre elles, généralement européenne, a apporté le lexique. Le créole est souvent resté au contact de sa langue lexicatrice ce qui est source d'une grande variation.

Le créole parlé en Guinée Bissau et en Casamance est né du contact entre portugais et langues africaines. En Guinée, il est resté non seulement au contact du portugais mais aussi des langues africaines. En Casamance, depuis 1886, le contact avec le portugais s'est perdu mais de nouveaux contacts avec le français et le wolof ont transformé la langue.



EVOLUTIONS ACTUELLES

Guinée Bissau, pourcentage de la population qui parle kriol :
1979 = 44,3% ; 2009 = 84,11%
Casamance, nombre de locuteurs du kriol :
1963 = 57 500 ; 2012 = 20 000

Mais aussi :

- fortes migrations Guinée -> Casamance
- exode rural
- développement des moyens de communication (télévision, radio, ...) qui entraîne la diffusion du kriol

Cette situation engendre une multiplicité de façons de parler le kriol. Comment les catégoriser ?

Le corpus se compose de :

- Nombre d'enregistrements : 47
- Nombre d'heures : 27h
- Transcriptions réalisées : 6 heures 59 minutes 30 secondes sur le logiciel Transcriber - participation de locuteurs natifs à la transcription

Procédure pour la collecte de données déjà effectuée :

- entretiens semi-guidés
- répertoires linguistiques variés (monolingue, bilingue, plurilingue)
- situations "authentiques" (bruits, interférences de communication, ...)
- adultes hommes et femmes

LES DÉFIS DU TAL

- **Tout transcrire ?**
 - manuellement ? automatiquement ?
 - type d'orthographe ?
- **Redire ?** une alternative à la transcription, la méthode du respeaking proposé par le logiciel LigAikuma est-elle possible ?

Dans tous les cas, quel niveau de finesse adopter pour permettre de prendre en compte les particularités socio-linguistiques :

pode vs pudi ((il) peut)
meste vs misti (vouloir)

- **Que faire des innovations ?**

La méthode du repérage des frontières prosodiques pourrait-elle servir à décrypter s'il y a une ou deux langues (influence ou pas du portugais ancré dans la langue) à l'aide des logiciels JTRANS (découpage en tokens) et SLAM (catégorisation prosodique).

voyagem = voyage (français) + -agem (portugais)
-> en kriol *byas*

```

EN
+ [pron=pi] Ami N na buri mas , si e pudi
ma fasilita nu voyagem (cd mixt. français/portugais) na kamiñu purke
ami N pensa kuma diferu di Senegal
partiku ...
partikularmentiiii
li na ziquintxor anos i familia
ma kada frontera ku bu txiga kada kontrolu ku bu txiga anos koytadi e ta pudu pagamentu [pron=pi]
e ta pudu pagamentu son
agora si bu say li
talbes bu ta say li i ka ka sabi i ka sabi sabura ki bu leba-u

```

RESSOURCES ACTUELLES ET À VENIR

Création d'un site web dédié au projet de recherche sur cette langue



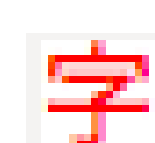
```

TINO
[i] Naw na bardadii
amiii n padidu na bisaw[i]
naa bayru deë ganbyafada
na anu di mil nofsentus i sinkuenta i oytu
eëë ña pape ta txomadu duarti da silva ña mame ta txomadu Maria da silva
i ma pa n fala so kumaa tend e konta+ [lang=Portuguese] di ki
mumentu [lang=Portuguese] ku n padidu yara mumentu nasionalista na ginebisaw
na gine portuguesa su{si bu} misti
ña pape ku ña mame sta ba obrigadu dii refujia+ [lang=Portuguese] naa senegal

```



Dépôt du corpus oral sur la plateforme COCOON



Construction collaborative d'un dictionnaire kriol/français sur le modèle du projet Jibiki