

## Description et Analyse des Méthodes de Traitement Automatique de la Phonologie Dugun à Partir de Phonology Assistant

Albert Blangar\*

### Résumé

Cet article vise à explorer l'apport des outils du Traitement automatique des langues naturelles (TALN), notamment le logiciel Phonology Assistant (PA), appliqué à l'étude phonologique du dugun. La problématique est centrée sur l'utilisation efficace des outils d'analyse automatique pour la description des langues africaines faiblement dotées. L'exploitation des outils d'analyse automatique contribue significativement à l'acquisition des compétences dans le cadre de la formation en linguistique africaine, telle est notre hypothèse. L'approche utilisée est le traitement automatique du langage naturel (TALN). L'étude s'appuie sur la méthode mixte et l'analyse de contenus. Le corpus se compose de deux volets: un ensemble de données linguistiques comprenant des captures d'écran d'interfaces des outils du TAL et une liste de 1 700 mots transcrits phonétiquement, ainsi qu'un ensemble documentaire constitué des programmes académiques. Globalement à l'issue de nos analyses, l'application de Phonology Assistant à l'analyse phonologique de la langue dugun au Cameroun, débouche sur deux résultats fondamentaux: la méthode d'analyse automatique du dugun, avec la nécessité de combiner plusieurs outils du TAL pour des analyses adéquates et satisfaisantes, mais aussi de vulgariser ces outils numériques dans les universités au Cameroun à partir de l'implémentation d'un parcours professionnel en ingénierie linguistique.

**Mots clés :** Analyse, TAL, phonologie, dugun

### Description and Analysis of Automatic Processing Methods for Dugun Phonology Using Phonology Assistant

#### Abstract

This paper aims to explore the contribution of Natural Language Processing (NLP) tools, particularly the software Phonology Assistant (PA), to the phonological study of Dugun. The central research question focuses on the efficient use of automatic analysis tools for the description of African languages. We hypothesize that leveraging automatic analysis tools significantly enhances skill acquisition within the framework of African linguistics training. The approach employed is natural language processing. The study is based on a mixed-methods design and content analysis. The corpus consists of two components: a linguistic dataset including screenshots of NLP tool interfaces and a list of 1,700 phonetically transcribed words, as well as a documentary corpus composed of academic programs. Overall, our analysis shows that the application of Phonology Assistant to the phonological analysis of the Dugun language in Cameroon yields two main outcomes: first, the automatic analysis method of the dugun, of the need to combine multiple NLP tools for adequate and satisfactory analyses, and second, the need to popularize these digital tools in Cameroonian universities through the implementation of a professional curriculum in linguistic engineering.

**Keywords:** Analysis, NLP, phonology, dugun

\* Université de Ngaoundéré Cameroun, Faculté des Arts Lettres et Sciences Humaines, Département de Langues Africaines et Linguistique, Laboratoire Langues Dynamiques et Usages(LADYRUS), Ngaoundéré, Cameroon albertblangar@gmail.com, Orcid :0009-0008-6141-094X

**Reçu (Received):** 25.09.2025

**Accepté (Accepted):** 06.01.2026

**Publié (Published):** 29.03.2026

**Citation/Cite:** Blangar. A. (2026). Description et Analyse des Méthodes de Traitement Automatique de la Phonologie Dugun à Partir de Phonology Assistant. *ChronAfrica*, 3(1), 74-98. <https://doi.org/10.5281/zenodo.19233068>

Copyright © 2026 The Author(s). This is an open-access article under the Creative Commons Attribution License (CC BY) which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is properly cited.

## Introduction

Dans un contexte où les autorités locales au Cameroun valorisent la professionnalisation des filières dites « facultaires » et promeuvent la digitalisation dans les secteurs éducatifs, nous constatons sur le plan didactique, l'usage inapproprié et peu efficace des méthodes d'analyse automatique par les enseignants et les étudiants. Sur le plan de la recherche, on note un faible recours aux outils du Traitement automatique de langue (TAL) pour la description des langues camerounaises. La problématique de l'usage adéquate des outils d'analyse automatique des langues africaines et l'innovation pédagogique dans les universités d'État au Cameroun est au centre de notre réflexion. Selon Mboning et Wandji (2021), la plupart des outils du TAL reste concentrée sur les langues riches. Mais nous pensons qu'il existe de nos jours des outils adaptés aux langues à faible ressources. C'est pourquoi nous nous intéressons à l'expérimentation du logiciel *Phonology Assistant* appliqué au dugun, une langue camerounaise insuffisamment décrite, afin d'en proposer une étude phonologique. Dans quelle mesure le recours aux outils de TAL facilite-t-il la description, l'analyse phonétique et phonologique d'une langue à tradition orale telle que le *dugun* ? De cette question, se dégage hypothèse selon laquelle, l'intégration des méthodes de TAL améliore la fiabilité et l'automatisation de la description linguistique en facilitant l'acquisition des compétences et la documentation. Notre objectif est d'analyser les potentialités offertes par *Phonology Assistant* combiné à d'autres outils du TAL. L'article est structuré en trois points majeurs : l'ancrage théorique et méthodologique, les résultats de l'expérimentation de *Phonology Assistant* (PA), les défis et réajustements relatifs à l'assistant (PA).

### 1. L'outil et la langue d'expérimentation

Avant de développer les points majeurs de notre travail, nous présentons brièvement le logiciel d'étude et la langue d'application.

*Phonology Assistant* (PA), assistance phonologique en français, est un logiciel du Traitement automatique du langage naturel (TALN) conçu par la Société internationale de linguistique (SIL) permettant aux linguistes d'analyser les données phonétiques et phonologiques. Cet outil a été mis en ligne pour la première fois en 2008 (version 3.0.1). L'assistant (PA) permet d'identifier, décrire et d'analyser automatiquement des phénomènes phonologiques, proposer des inventaires phoniques des consonnes, des voyelles, le classement des matériaux (tableaux phoniques), l'identification des phonèmes (des paires minimales, environnement identique...). L'outil explore les données phonétiques et recherche des preuves de contraste ou de distribution complémentaires.

Le logiciel n'a pas variablement fait l'objet de plusieurs travaux scientifiques surtout au Cameroun. Toutefois, hors mis les travaux effectués par des linguistes de la Société internationale de linguistique (SIL), nous notons les plus importants réalisés par Mirdom Teronpi (2005), dans son ouvrage intitulé *A Preliminary Description of the Phonology of Ri-Bhoi Amri Karbi*. L'auteur indique que les données ont été analysées grâce à l'utilisation de l'outil *Phonology Assistant*. En outre, les travaux réalisés par Timothy Kempton & Roger K. Moore (2009), *Finding Allophones : an Evaluation on Consonants in the TIMIT Corpus*. Les auteurs disent explicitement avoir converti les transcriptions en Unicode Alphabet phonétique international (API) en utilisant *Phonology Assistant*. Mark Dingemans quant à lui, présente dans la revue technique de *Phonology Assistant* (2008) l'utilisation de cet outil en linguistique



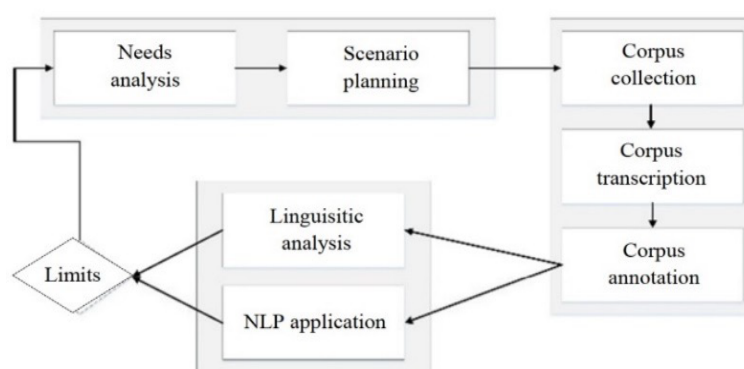
mixte qui combine des analyses quantitatives et qualitatives, et l'analyse de contenus. La méthode mixte, Dörnyei (2007) et McEnery & Hardie (2012) est idéale pour étudier les langues à partir de corpus, car elle relie l'analyse linguistique et les traitements automatiques pour une analyse plus riche et équilibrée d'un phénomène linguistique ou didactique. Nous prenons en compte les données qui relèvent du traitement automatique de la langue (analyse linguistique) d'une part, et les données issues des entretiens d'autre part. Pour la première catégorie des données, l'outil utilisé est l'analyse quantitative (traitement automatique de la langue *dugun* et questionnaire) et/ou qualitative (entretien). L'étude est construite sur les résultats d'une enquête sur les pratiques enseignantes. En ce qui concerne la seconde catégorie de données, l'outil d'analyse utilisé est l'analyse de contenu (recension des documents, description, classification, interprétation). Ici, il s'agit, entre autres, des programmes d'enseignement, syllabus, fiches de suivi des enseignements, des tutoriels existants.

## 2.1 Collectes de données

Nous avons deux types de corpus. Les données linguistiques et l'analyse documentaire. Le premier type est constitué des données linguistiques servant à l'exemplification. Il s'agit des images issues de captures d'écran des interfaces des outils du TAL à étudier et la liste de mille sept cents mots (1 700) transcrits phonétiquement destinée à la description, à l'analyse phonologique et tonologique. Le second type est l'analyse documentaire, il est constitué des programmes d'enseignement académique, des emplois de temps. Pour la collecte et l'analyse de nos données, nous nous sommes inspiré du schéma de cycle de vie d'un corpus selon Mohamed, Z. K. (2016), (voir figure 1).

### Figure 1.

Schéma d'un cycle de vie d'un corpus assisté par l'outil du TALN



Source. Adapté du schéma d'un cycle de vie d'un corpus par (Mohamed, 2016).

Premièrement, nous avons procédé à l'identification des besoins. Pour ce fait, nous nous sommes rendu à Gouna, chez les *Dugun* pour le choix de dialecte de référence. Deuxièmement, avons défini le contexte de travail (en communauté), à partir d'une base de données de 1 700 mots, corpus que nous avons enregistré au moyen d'un téléphone smartphone Camon17, en collaboration avec les locuteurs natifs. Cet atelier s'est passé en une semaine. Ce fut en février 2021. Troisièmement, sous la base de la liste de 1 700 mots que nous avons définie, nous avons commencé avec la collecte de données proprement dite.

Quatrièmement, la transcription de notre corpus. Nous avons entièrement transcrit la base de données de 1 700 mots en version manuscrite.

## **2.2 Annotation du corpus et analyse de données linguistiques**

La phase d'annotation du corpus consistait à segmenter les tons, les catégories des mots, les variantes plurielles etc. Ensuite, nous avons procédé à la phase d'analyse linguistique en laboratoire, souvent avec l'aide des locuteurs *dugun*. Nous avons fait usage des outils TAL tels que *FLEx*<sup>1</sup> (*Field Language Explorer*), l'outil (PA) et *dekereke* (caméléon en français). Dans notre étude, nous avons travaillé avec des femmes et hommes âgées et ayant une bonne dentition, bon appareil phonatoire, des locuteurs natifs du village. Ces exigences nous a permis de mieux proposer des inventaires phoniques exactes pour une analyse efficiente des phénomènes phonologiques facilement identifiables par les outils du TAL.

## **2.3 Annotation du corpus et analyse de données linguistiques**

Nous avons organisé un atelier de confrontation de données qui s'est déroulé en communauté pour la vérification et l'évaluation des données et nos analyses faites en laboratoire.

## **3. Les résultats de l'expérimentation de Phonology Assistant pour l'étude phonétique et phonologique du dugun.**

Le résultat s'articule autour de trois points majeurs : les méthodes d'utilisation de l'assistant (PA) pour l'analyse phonologique, son interopérabilité avec le logiciel *Field Language Explorer (FLEx)* et *dekereke* sa limite et réajustement.

Essono (2006) présente cinq branches de la phonétique notamment, la phonétique synchronique, la phonétique diachronique, la phonétique appliquée, la phonétique expérimentale et la phonétique générale. En effet, la phonétique appliquée regroupe en son sein : la technologie des télécommunications, la digitalisation de la parole, l'utilisation des structures acoustiques des sons, etc. La phonétique expérimentale ou instrumentale quant à elle s'intéresse à l'étude des messages vocaux (émis ou reçu) à l'aide d'instruments servant à l'analyse et à l'enregistrement des signes acoustiques. Nous avons spécifiquement choisi la branche de la phonétique appliquée et expérimentale pour nous permettre d'étudier les outils du TALN appliqués à la phonologie du *dugun* via l'assistant (PA). Concrètement comment l'outil (PA) peut-il aider à effectuer l'analyse phonétique et phonologique : identification des inventaires phoniques des consonnes, voyelles et d'autre phénomènes tels que les paires minimales, les paires suspectes, des environnement identiques, le profile syllabique, le marquage des tons, etc.?

### **3.1 Méthode de transcription, de configuration et d'importation de 1 700 mots**

Dans cette session, nous allons décrire et explorer les méthodes qui conduisent à la transcription

---

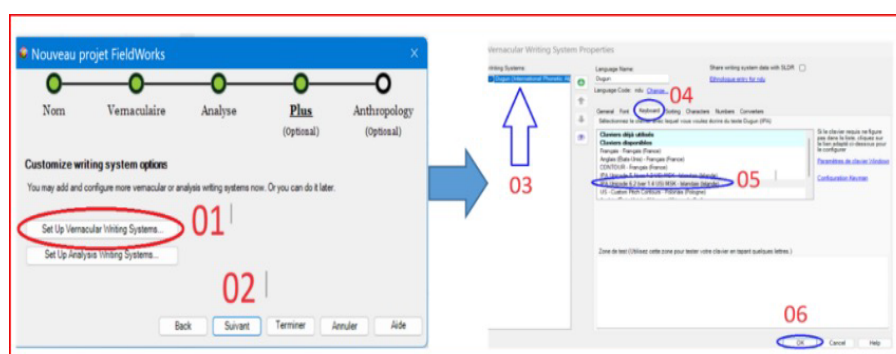
<sup>1</sup> En français FLEx signifie champ d'exploration linguistique.

phonétique pour l'analyse phonologique du dugun à travers l'assistant (PA). Cependant, avant d'effectuer cette tâche de transcription phonétique, il est important de télécharger et d'installer au préalable des outils tels que *FLEX* et l'assistant (PA). Pour télécharger gratuitement ces outils, nous nous sommes rendus sur le site de l'organisation SIL(2012) à travers le lien [www.silcam.org](http://www.silcam.org). À ce stade, nous pouvons avoir accès aux tutoriels facilitant l'installation et le démarrage de ces outils. Après avoir installé *FLEX* sur notre ordinateur, nous avons créé un projet nommé (Phonologie dugun). Puis, nous avons commencé à transcrire la base lexicale de 1 700 mots à partir de *FLEX*. Après avoir mené l'activité de transcription, l'assistant (PA) récupère automatiquement les informations contenues dans *FLEX* tel que décrit dans les figures 2,3,4 et 5. La transcription phonétique des mots est faite sous le modèle d'Alphabet phonétique international(API).

Après avoir installé *FLEX* dans l'ordinateur, il est également important d'installer un clavier qui prend en charge les caractères des langues africaines ; Galicien, IPA Unicode 5.1 (ver 1.2 US) MSK, IPA Unicode 6.2 (ver 1.4 US) MSK, Cameroun AZERTY MSKLC Unifie 2011, Islandais, etc., ils sont également à libre accès. Pendant l'installation, *FLEX* nous demande de configurer un système d'écriture. Deux options sont possibles, le système d'écriture d'analyse » et le système d'écriture vernaculaire ». Dans le cadre de notre projet, nous avons choisi le «système d'écriture vernaculaire ». La figure 2 ci-dessous résume les étapes de configuration du système d'écriture. L'ordre est indiqué par la numérotation allant du numéro 01 au numéro 06.

**Figure 2.**

*Schéma de configuration du système d'écriture des langues africaines (FLEX)*



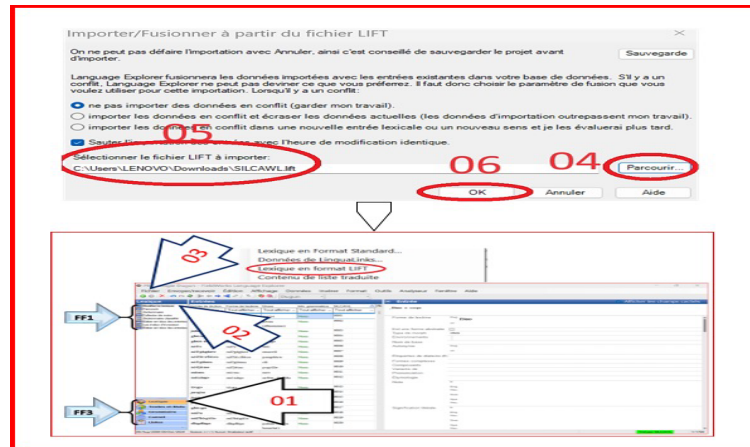
Source : Créée par l'auteur.

Après avoir installé le clavier sur l'ordinateur, nous avons choisi ensuite la liste de 1 700 mots, SIL Comparative African Word List (SILCAWL) en anglais, c'est un format standard élaboré par Keith et Roberts (2004) afin de résoudre un problème spécifique relatif à la standardisation du système d'écriture des langues africaines. La liste de 1 700 mots contient les informations grammaticales, avec ses variantes plurielles, des expressions anthroponymiques, toponymiques et zonymiques, les parties physiques et non physiques des êtres humains, les noms au sujet des personnes, les activités et les civilisations humaines, les noms des animaux, les plantes, l'environnement, les événements et les actions, les qualités, les quantités, les items grammaticaux, etc. Après avoir transcrit phonétiquement la liste de 1700 mots, l'étape suivante consiste à importer la base des données des mots transcrits dans le logiciel assistant (PA). Comment cela fonctionne-t-il de manière pratique ?

Nous proposons aux utilisateurs de suivre les étapes que nous avons suivies telles que représentées par les figures 3, 4 et 5 ci-dessous.

### Figure 3.

Schéma des étapes vers l'importation de la liste de 1 700 mots vers FLEx



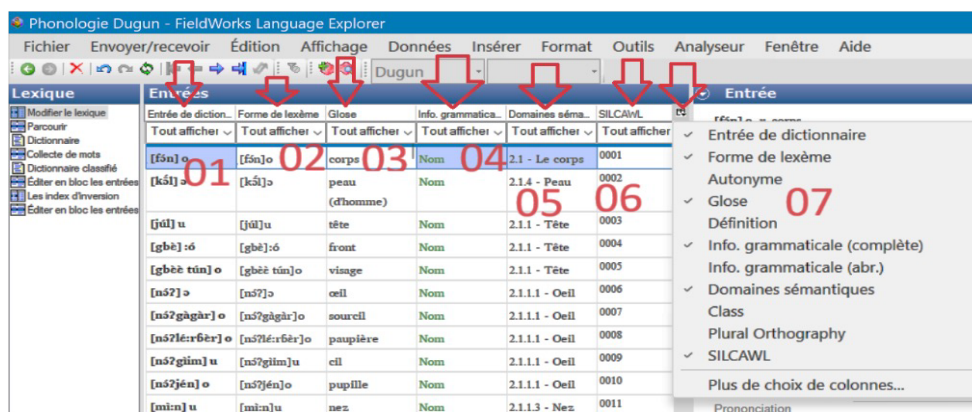
Source : Créée par l'auteur.

Une fois l'interface de l'outil *FLEx* ouverte, nous avons cliqué sur l'onglet « Lexique » dans la « zone de sélection (FF3, (01), ensuite sur la fenêtre « Modifier le lexique » dans la « sélection d'Affichage(FF1) », numéro 02. Puis, sur « Fichier », (03), nous avons choisi l'onglet « Lexique en format LIFT » pour l'importer dans l'outil *FLEx* en effectuant une clique sur la fenêtre « Parcourir » (04). L'étape cinq (05) consiste à sélectionner le format LIFT sauvegardé dans un dossier de notre ordinateur. Arrivé à ce niveau, il ne reste qu'attendre la réaction finale de l'ordinateur, c'est-à-dire l'importation complète du fichier LIFT. Pour finaliser le processus d'importation, nous avons cliqué sur « ok » (étape 06).

L'étape sept (07), figure 4 ci-dessous indique la zone de paramétrage des entrées pour rendre certaines fonctionnalités visibles, notamment l'affichage des domaines sémantiques, les formes de lexème, l'information grammaticale, les gloses, etc. L'entrée de la colonne 01 est réservée à l'édition des lexiques. La colonne 02 est la forme lexicale modifiable dans l'outil *FLEx*. La colonne 03 est consacrée aux gloses en français. Le numéro 04 concerne les informations grammaticales (catégories). La colonne 05 est la structuration de la liste de 1700 mots en sous domaines sémantiques. Le 06 est la numérotation des autonymes. Et enfin la dernière colonne est la fenêtre de paramétrage des entrées masquées. Cette transcription phonétique respecte la rigueur scientifique relative aux normes édictées par l'API, déjà intégrées dans le logiciel *FLEx*.

**Figure 4.**

*Schéma de l'interface de la liste de 1 700 mots*



*Source : Créée par l'auteur.*

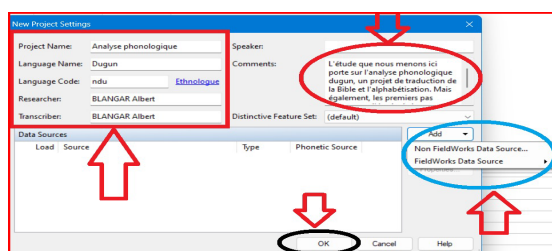
De ce qui précède, nous retenons que la méthode de transcription phonétique, de configuration et d'importation de la liste de 1 700 mots pour l'analyse phonologique du *dugun* à partir de l'assistant (PA) nécessite au préalable que l'on télécharge et installe le clavier des langues africaines, le logiciel *FLEX* tout en y créant un projet. Ensuite, procéder à l'importation de la base de 1 700 mots, entièrement transcrit.

### 3.2 Méthodes de création d'un projet dans l'outil assistant (PA)

Dans cette sous-session, nous allons procéder à quelques techniques de création d'un projet dans le logiciel PA et montrer comment explorer les inventaires phoniques. Pour créer le projet d'analyse phonologique du *dugun* dans l'outil (PA), nous avons suivi quatre étapes suivantes : le Renseignement de la bande carrée en rouge consacrée aux métadonnées (voir figure 5). Le recours au cercle en rouge, réservé au commentaire ou à la description du projet. Celui en bleu est une entrée qui oriente l'utilisateur vers les données sources contenues (liste de 1 700 mots) transcrit phonétiquement dans *FLEX*. Et le cercle en noir est la dernière étape de la configuration (création d'un nouveau projet). Une Clique sur « Ok », l'outil (PA) importe automatiquement la base de données de 1 700 mots. Ces étapes sont illustrées par la figure 5 ci-dessous.

**Figure 5.**

*Schéma des étapes vers l'importation de la liste de 1 700 mots vers FLEX*



*Source : Créée par l'auteur.*

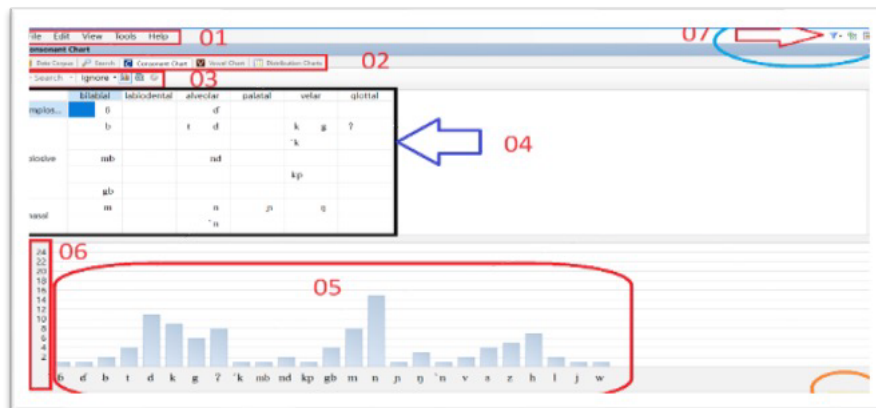
### 3.3 Méthodes d'exploration et distribution des caractères phoniques

Ici, nous allons décrire les techniques de consultation des inventaires des consonnes et voyelles dans l'outil (PA).

Après avoir créé un projet d'étude, l'outil récupère systématiquement les informations contenues dans l'outil *FLEX* et procède à la génération des inventaires des consonnes et voyelles, sur la base du corpus. La figure 6 ci-dessous est un cas illustratif des inventaires consonantiques. Il est constitué principalement de sept (07) onglets : l'onglet « Fichier », « Éditer », « Affichage », « Outils », « Aide », « Bande de corpus », « Bande recherche », « Tableau des consonnes », « Tableau des voyelles » et « Tableau de répartition phonique ». Cet inventaire prend en compte les lieux d'articulation, le mode d'articulation, des caractères digraphes, les occurrences des sons, etc. L'inventaire phonique proposé présente une fréquence selon leur positionnement dans les mots (allant des caractères les plus utilisés vers les moins utilisés). Ainsi donc, l'inventaire des consonnes dugun est identifiable sur la figure 6 par le cadre en noir (numéro 04). La bande en rouge (05 et 06) est le regroupement des sons consonantiques *dugun*. On peut remarquer que le son [n] est le plus utilisé dans la langue et les moins utilisés sont entre autre les sons [j] et [w]. Lorsqu'on clique sur chaque son ou phonème, l'assistant (PA) nous oriente immédiatement vers les exemples des mots où la lettre cible est utilisée, ceci selon leur positionnement dans les mots (initial, médiane et final). La figure 6 ci-dessous est l'inventaire des consonnes *dugun* obtenu de manière automatique.

**Figure 6.**

*Schéma des inventaires phoniques des consonnes à l'aide de de l'outil (PA)*



Source : Créée par l'auteur.

Pour accéder à la statistique des caractères phoniques contenus dans les mots, on peut consulter chaque son dans le cadre en rouge (voir figure 6). Pour obtenir les [digraphes], une configuration préalable est requise, faire usage de la fenêtre nommée séquence ambiguë « Ambiguous sequence », en paramétrant les informations telles que l'on peut le constater sur la figure 7 suivante.

Pour obtenir l'inventaire phonique des voyelles, nous avons procédé à la même méthode que précédente (voir figure 6) et choisir rubrique voyelle.

**Figure 7.**

*Schéma des inventaires phoniques des consonnes à l'aide de de l'outil (PA)*

Sequence	Treat as one unit?	Base Character	CV Pattern
kp	<input checked="" type="checkbox"/>	k	C
gb	<input checked="" type="checkbox"/>	b	C
nd	<input checked="" type="checkbox"/>	d	C
mb	<input checked="" type="checkbox"/>	b	C
˘ŋ	<input checked="" type="checkbox"/>	ŋ	C
˘[	<input checked="" type="checkbox"/>	[	[
˘u	<input checked="" type="checkbox"/>	u	V

*Source : Créée par l'auteur.*

En plus des inventaires des consonnes et voyelles, l'assistant (PA) offre la possibilité de visualiser les caractères phoniques selon leur fréquence d'emploi ou mode de distribution dans les mots. Les figures (8 et 9) nous présentent la méthode de « distribution des caractères » phoniques des consonnes (voir figure 8) et des voyelles (voir figure 9). La distribution phonique des caractères prend en compte la position initiale, médiane et finale et leur mode d'articulation.

Pour mieux faciliter la compréhension, nous allons décrire ce tableau en expliquant chaque entrée telle que décrit sur la figure 8.

**Figure 8.**

*Schéma d'un tableau de distribution des consonnes selon leur occurrence*

Chart Name: ExConsonants	01	02	03	04	05	06	07	08	09
<#	6	109	26	94	15	6	0	25	78
*[V]	28	372	212	177	181	4	10	193	140
[V]*	6	69	25	56	14	6	0	24	43
#_*	11	119	61	70	50	1	0	55	57
[V][...]	15	120	90	39	72	3	5	84	29
*_[C]	0	15	15	0	15	0	0	13	0
[C]*	0	40	1	38	1	0	0	1	35
[C][V]*	0	23	0	23	0	0	0	0	18
*_[V][C]	2	85	46	36	44	0	5	41	28
[[C][voiceless]]	0	35	1	34	1	0	0	1	29
[[C][voiced]]	0	0	0	0	0	0	0	0	0
[[plosive][voiceless]]	2	14	19	1	14	4	0	19	1
[[plosive][voiced]]	0	14	0	13	0	0	0	0	13
[[C][nasal]]	51	14	17	4	5	91	135	15	0
{[trill],[tap],[flap]}	54	0	5	1	0	4	52	5	0
[[fricative][voiceless]]									
[[fricative][voiced]]									
[approximant]									
[+strid]									
[[C][◌ <sup>h</sup> ]]									
?									
[+labiodental]									
[high tone]									
[mid tone]									

*Source : Créée par l'auteur.*

La figure 8 ci-dessus comporte trois parties : un cadre en bleu, un cercle rouge situé au-dessus, marquée par la flèche en rouge et numérotés de 01 à 09. En fin la partie en jaune marquée par « 0 ». La bande contenant les cercles numérotés de 01-09 est la statistique de des sons (consonnes) dans les mots ou syllabes, suivant le mode d'articulation. Par exemple, les symboles [ \_#] signifie que l'outil assistant (PA) a inventorié et classifié toutes les consonnes selon leur mode d'articulation (sonores, sourdes, nasales, fricatives, etc.) en position finale dans les mots.

[#\_ ] → inventaire des sons où le caractère (consonnes) occupe la position initiale ;

[ \_#] → inventaire des sons où le caractère occupe la position finale ;

[V]\_ ] → inventaire des sons où le caractère est en intervocaliques ou médianes, etc.

Il est important de noter que les carreaux en jaune nous indiquent que l'outil (PA) n'a pas pu identifier des sons appartenant à cette catégorie phonique ou tout simplement ces sons n'existent pas dans la langue d'étude. Ces carreaux sont représentés par des chiffres « 0 » (voir figure 8 et 9). Le cadre en bleu, indiqué par une flèche (voir figure 8) est le mode d'articulation des caractères phoniques opéré et classifié automatiquement par le logiciel (PA). Lorsqu'on effectue des opérations en cliquant sur chaque numéro dans le carreau en bleu, on peut avoir accès aux différents sons ou phonèmes tels qu'ils apparaissent dans les mots et selon leur mode d'articulation et leur zone d'articulation. Ces modes d'articulations sont entre autre :

Les consonnes sourdes, [[C][Voiceless]]	Les vibrants {[Trill],[tap],[flap]}
Les consonnes sonores, [[C][Voiced]]	Fricative sourdes, [[Fricative][voiceless]]
Les occlusives sourdes, [[Plosive][voiceless]]	Fricative sonores, [Fricative][voiced]]
Les occlusives sonores, [[Plosive][voiced]]	Approximant
Les consonnes nasales, [[C][Nasal]]	[Les labiodentales]

En ce qui concerne la distribution des voyelles, nous nous referons à la statistique présentée par la figure 9 ci-dessous. Cet ensemble des sons vocaliques *dugun* est généré automatiquement grâce à l'assistant (PA).

**Figure 9.**

*Schéma d'un tableau de distribution des consonnes selon leur occurrence*

	[C]_#	#_*	*_*	[V]_*	*_[V]	*_[approxima...]	[approxima...]
L.01	[V]	217	16	536	38	38	46
L.02	[[V][ + high]]	48	7	120	9	10	10
L.03	[[V][ -high][ -lo...]]	156	7	325	27	20	28
L.04	[[V][ + low]]	13	2	91	2	8	10
L.05	[[V][ + rnd]]	179	9	324	31	16	20
L.06	[[V][ -rnd]]	38	7	212	7	22	26
L.07	[[V][ + front]]	38	7	198	7	21	24
L.08	[[V][ -front][ -b...]]	0	1	15	0	2	2
L.09	[[V][ + back]]	179	8	323	31	15	20
L.10	[long]	8	1	80	16	17	5
L.11	[high tone]	42	4	156	17	14	14
L.12	[low tone]	21	3	117	16	19	19
L.13	[mid tone]	50	1	58	5	0	1
L.14	[rising tone]	0	2	8	0	0	1
L.15	[falling tone]	0	0	5	1	2	0
L.16	[ + ATR]	153	11	308	21	18	27
L.17	[ - ATR]	64	5	228	17	20	19

Source : Créée par l'auteur.

La figure 9 ci-dessus, présente l'ensemble des inventaires vocaliques, représenté selon leur apparition dans les mots ou dans les syllabes. Les sons vocaliques sont repartis dans le tableau suivant leur fréquence d'emploi, leur mode d'articulation et leur position dans les mots. La grille encadrée et indiquée par la flèche (voir figure 9) en bleu représente la distribution des sons ou des phonèmes, selon leur positionnement dans les mots ou syllabe. La ligne (flèche) nommée L.02-L.09 est un regroupement. La ligne nommée L.10 est la fréquence phonique des voyelles longues. La ligne 11 jusqu'à la ligne 15 (voir L.11-L.15) représentent l'illustration des sons vocaliques ayant les tons : Ton Haut (TH), indiqué par L.11, le Ton Bas (TB), indiqué par L.12, Ton Moyen (TM), indiqué par L.13 et les tons modulés (BH et HB) sont indiqués par L.14 et L.15. Les deux dernières lignes (voir L.16 et L.17, figure 9) sont des ATR+ et ATR-.

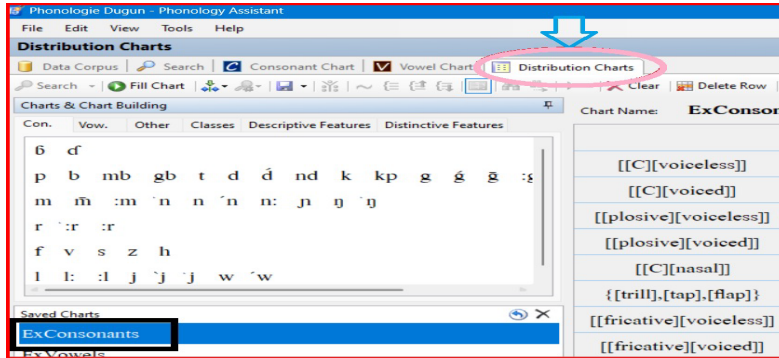
En conclusion, la création d'un projet grâce à l'assistant (PA) prend en compte les informations de la métadonnée et le code Iso de la langue d'étude. Phonology Assistant, combiné au logiciel FLEX nous a permis d'obtenir les inventaires phoniques des consonnes et des voyelles dugun de manière provisoire. L'outil a proposé également les fréquences d'emploi des sons selon leur position dans les mots. Par ailleurs, pour une analyse adéquate et satisfaisante, nous interpelons les utilisateurs de l'outil l'assistant (PA) à faire usage de l'approche hybride, une méthode combinant annotations manuelles et neuronale, Bird (2009).

### 3.4 Méthode d'exploration des paires minimales et environnements identiques

Notre but consiste à décrire les étapes d'exploration des tableaux de paires minimales (PM) et les tables des environnements identiques (EI) des consonnes et voyelles. La figure 10 ci-dessous est une interface qui montre les étapes de génération automatique des PM.

La première étape : Une clique sur l'onglet « Distribution chart », tel que représenté par le cercle en rose (flèche en bleu).

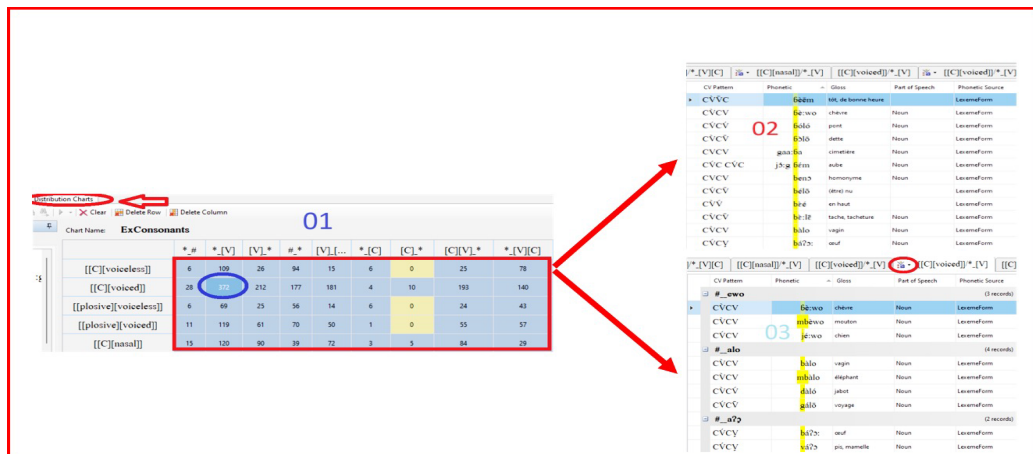
**Figure 10.**  
Onglet vers la génération de paire minimale



Source : Créée par l'auteur.

Après avoir choisi l'onglet nommé « exemple de consonnes » (cadre en noir voir figure 10 ci-dessus), nous avons cliqué sur l'onglet « Ex Consonant » (cadre en noir figure 9 qui veut simplement dire exemple des consonnes). Après cela, la fréquence des caractères s'affiche telle qu'on peut l'observer dans le cadre en rouge (voir figure 11). On a ensuite cliqué sur le petit cercle en bleu à l'intérieur du cadre. Des exemples de mots s'affichent tels qu'on peut voir dans la (voir figure 11).

**Figure 11.**  
Affichage de la paire minimale dans l'assistant (PA)



Source : Créée par l'auteur.

Dans cette partie, nous avons choisi les paires minimales (PM) avec les exemples des consonnes occlusives ([[C][voiced]]) (voir lettre en jaune, tableau 1,2,3,4,5,6,7). Ces PM sont indiquées par la première flèche en rouge montante vers (voir figure 11, étape 2). L'assistant (PA) a proposé tous les sons consonantiques (occlusives) du dugun avec leur occurrence s'affichant dans ce cadre (voir étape 2 et 3). Une fois cette étape terminée, la dernière étape consiste à créer un tableau de PM. Pour ce fait, nous avons cliqué sur le petit symbole (figure 10 étape 03). Une fois cliqué sur ce symbole, l'outil propose

alors le tableau de PM tel que représenté par l'étape 03. L'outil (PA) organise le tableau de PM selon le profil syllabique, la forme lexicale, la catégorie grammaticale et selon le mode d'articulation, position des caractères dans les mots, etc.).

En guise d'exemples d'application, (voir exemples 1, 2, 3, 4, 5, 6, 7, 8 et les tableaux 1, 2, 3, 4, 5, 6, 7) ci-dessous. Ces tableaux nous présentent les paires minimales des consonantiques voisées et non voisées, les environnements identiques dugun. Les PM peuvent apparaître en initiale, en médiane tout comme en final des mots. Il faut noter que l'assistant (PA) propose non seulement le tableau des PM, mais cet outil propose également les informations de la métadonnée : le titre du projet et le nom de la langue cible, le mode d'articulation, le nombre de mots enregistrés dans la table de PM, la date et l'heure d'extraction de données en laboratoire. Toutefois, il est recommandé que l'utilisateur choisisse la version finale de PM après avoir défini les critères d'identification de véritables PM. Selon Snider (2023), les véritables PM doivent avoir le même profil syllabique, la même phonation, même catégorie grammaticale, etc. Ci-dessous quelques exemples de PM du dugun générées par l'assistance de l'outil assistant (PA). Ces exemples sont numérotés de 1 à 8, avec les illustrations dans les tableaux ci-dessous. Exemple 1. Les paires minimales qui s'opposent en position initiale : [b-mb], [mb- b]...

Pendant l'exportation d'une fiche, l'outil propose automatiquement les informations telles que le mode d'articulation, l'occurrence des lettres identifiée (Number record), le nom ou le titre du projet, le nom de la langue d'étude, la date et l'heure d'extraction de données.

### Tableau 1.

*Illustration des paires minimales s'opposent en initial*

**Number of Records:**102

**Project:***Phonologie Dugun*

**Language:***Dugun*

**Date/Time:** 05/11/2024 14:10:57

**Search Pattern:** [[C][voiced]]/\*\_[V] (Minimal Pairs)

CV Pattern	Phonetic	Gloss	Part of Speech	Phonetic Source
CVCV	ʃè:wo	chèvre	Noun	LexemeForm
CVCV	mbèwo	mouton	Noun	LexemeForm
CVCV	bàlo	vagin	Noun	LexemeForm
CVCV	mbàlo	éléphant	Noun	LexemeForm

**Exemple 2.** Les paires minimales qui s'opposent en médianes/ [m- j], etc.

**Tableau 2.**

*Illustration des paires minimales s'opposent en médiane*

CVCV	sé mo	honte	Noun	LexemeForm
CVCV	sè jō	guerre	Noun	LexemeForm

**Exemple 3.** Paires minimales avec les consonnes occlusives sourdes

**Tableau 3.**

*Illustration des paires minimales avec des occlusives sourdes*

CVCV	Phonetic	Gloss	Part of Speech	Phonetic Source
CVCV	pɔ:ɡɔ	viande	Noun	LexemeForm
CVCV	tɔ:ɡɔ	oreille	Noun	LexemeForm

**Exemple 4.** Paires minimales avec les sons vocaliques en final

**Tableau 4.**

*Illustration des paires minimales où les sons vocaliques s'opposent en final*

CVCV	Phonetic	Gloss	Part of Speech	Phonetic Source
CVCV	kùmū	huile	Noun	LexemeForm
CVCV	kùnū	peau, pelure (de fruit)	Noun	LexemeForm

**Exemple 5.** Paires minimales avec les sons vocaliques en médiane [e-a], [e-ε], [a-o], [e-ɔ], [e-ɔ], [ε-ɔ], [e-a], etc.

**Tableau 5.**

*Illustration des paires minimales avec des son vocaliques s'opposent en médiane*

CVCV	hɛ:ɡò	bambou	Noun	LexemeForm
CVCV	ha:ɡó	corbeau	Noun	LexemeForm
CVCV	gb e:go	fœtus	Noun	LexemeForm
CVCV	gb ε:go	langue	Noun	LexemeForm

Il faut noter que nous également des cas des PM où il y a opposition en final avec des connes fricatives, des sons vocaliques en initiale

**Exemple 6.** Opposition en contexte identique avec des consonnes à l'initiale

**Tableau 6.**

*Illustration en contexte identique avec consonnes à l'initiale*

CVV	Phonetic	Gloss	Part of Speech	Phonetic Source
CVCV	pa:ɔ	oncle maternel	Noun	LexemeForm
CVCV	taámo	(être) fatigué	Qualificatif	LexemeForm

**Exemple 7.** Opposition en contexte identique avec des consonnes s'opposant en médiane ([S- ?])

**Tableau 7.**

*Illustration en contexte identique avec consonnes s'opposant en médiane*

CVCV	Phonetic	Gloss	Part of Speech	Phonetic Source
CVCV	ndà:sɔ:	bœuf châtré, bovillon	Noun	LexemeForm
CVCV	bá ?ɔ:	œuf	Noun	LexemeForm

**Note :** *Toutes ces consonnes sont des phonèmes distincts.*

**Exemple 8.** Opposition en contexte identique avec des voyelles s'opposant en initial ([i-ɔ], [a-ə], [o-ɔ])

En *dugun*, on retient que toutes les voyelles s'opposent en initial, en médiane et en final. Mais l'exemple suivant est un cas où des voyelles s'opposent uniquement en initial.

**Tableau 8.**

*Illustration en contexte identique avec des voyelles en initial*

CV Pattern	Phonetic	Gloss	Part of Speech	Phonetic Source
VCV	iru	deux	Numeral	LexemeForm
VCV	ɔɔ	mouton	Noun	LexemeForm
VCV	agɔ	danger	Noun	LexemeForm
VCV	əgɔ	sabot	Noun	LexemeForm

Au final, on note que toutes les consonnes et les voyelles du *dugun* sont des phonèmes distincts.

#### 4. Quelques limites et réajustements relatif à l'usage de *Phonology Assistant*

Dans cette partie, nous présentons quelques défis majeurs et pistes de réajustements en rapport avec *Phonology Assistant* et les utilisateurs. Les limites de l'outil (PA) sont situées à deux niveaux : au niveau de l'outil d'une part et au niveau des utilisateurs d'autre part.

##### 4.1 Limites au niveau de l'usage de *Phonology Assistant*

En ce qui concerne l'assistant (PA) proprement dit, malgré l'importance de ce logiciel, nous avons identifié quelques manquements. Primo, l'interface est essentiellement en anglais, espagnol et portugais. Secundo, la transcription phonétique est effectuée à travers un autre logiciel notamment *FLEx*, cela rend l'utilisation de l'assistant (PA) plus complexe. Tercio, le logiciel assistant PA ne prend pas en compte des enregistrements sonores, sauf partir d'un autre programme compatible comme Sound File Splitter. Pour finir, les caractères digraphes ne sont pas directement inventoriés dans le logiciel, cela nécessite une configuration au préalable.

Au niveau des utilisateurs de l'outil (Étudiants, enseignants et linguistes de terrain), quelques défis sont à signaler. En effet, À l'issue des questionnaires et entretiens menés auprès de nos informateurs, et sur la base d'analyse de contenus des documents institutionnels sur les pratiques enseignantes, l'interprétation de données révèle que sur 145 informateurs interrogés, 95 informateurs affirment que les pratiques de salles en lien avec les outils du TAL sont théoriques, soit 65,5%, 50 informateurs ont déclaré que les pratiques enseignantes sont à la fois théoriques et pratiques, soit un taux de 34%. Seulement 10/145 informateurs ont recours à l'utilisation des outils du TAL, soit 6,80% et 95/145 accordent une place marginale aux pratiques du traitement automatique des langues (TAL). En outre, on note l'utilisation inappropriée de certains outils qui ne prennent pas en compte la syntaxe et l'interface des langues africaines notamment le cas de « *IRaMuTeQ*<sup>2</sup>, le manque de formation pratique sur l'utilisation des outils du TAL disponibles et le laboratoire non équipé, parfois non fonctionnel.

##### 4.2 Limites au niveau de l'usage de *Phonology Assistant*

Pour une meilleure transcription phonétique et analyse des unités tonales, pour une meilleur élaboration des inventaires phoniques des consonnes, la génération des fréquences consonantiques, vocaliques, et structure syllabique nous proposons l'interopérabilité entre l'assistant PA, *FLEx* et *Dekereke*<sup>3</sup>. Dans cette partie, nous allons insister sur l'outil *dekereke* pour compléter les limites de l'assistant PA.

###### 4.2.1 Au niveau des inventaires phoniques des consonnes et leur fréquence

En observant la figure 12 ci-dessous généré par l'assistant PA, on remarque que la réalisation automatique des inventaires consonantiques en dugun n'est pas complète, certains sons n'apparaissent pas dans ce tableau, ou tout au moins cela nécessite une configuration très complexe. C'est le cas avec des digraphes [mb], [nd], [vb], etc.

<sup>2</sup> Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. En fait, le logiciel n'est qu'une interface d'une autre logiciel appelé R qui fournit des données statistiques. Il s'agit de textométrie.

<sup>3</sup> *Dekereke* est un outil d'analyse phonologique mis au point par Société internationale de linguistique.

**Figure 12.**

Interface de tableau des inventaires des sons consonantiques(dugun)

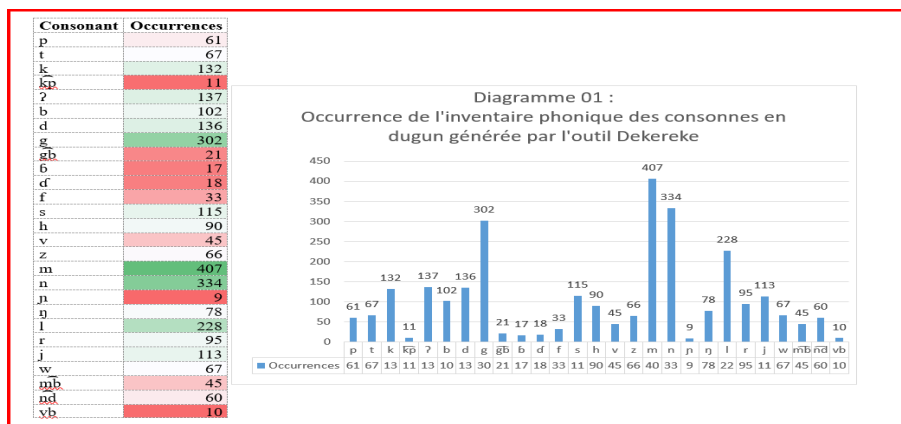
labial	Labiodental	Alveolar	Palatal	Velar	Labial-velar	Uvular	Laryngeal
p		t		k	kp	q	ʔ
b		d		g	gb		
ɓ		ɗ					
	f	s					h
	v	z					
m		n	ɲ	ŋ			
		l					
		r					
			j		w		

Source : Créée par l'auteur.

En plus, l'une des particularités de l'outil Dekereke est sa capacité à faciliter la construction des tables des fréquences suivies facilement exportable en diagramme. Une telle interface peut faciliter la tâche pour l'élaboration des syllabaires dans le champ d'alphabétisation(voir figure 13).

**Figure 13.**

Schéma de la Fréquences phonique des consonnes du dugun



Source : Créée par l'auteur.

Le tableau 9 ci-dessous est le résultat final des consonnes phonétiques dugun générés automatiquement puis arrangés manuellement grâce à *dekereke*. Sur la base de la liste de 1 700 mots transcrite phonétiquement, le *dugun* compte 32 sons consonantiques et non 25 sons tel qu'indiqué par l'assistant PA. Les détails sont mentionnés dans le tableau 9 suivant :

**Tableau 9.**

*Organisation manuelle des inventaires des sons consonantiques*

MODES D'ARTICULATION		POINTS D'ARTICULATION									
		Labiales		Alvéolaires			Palatales	vélares		glottales	
		Bilabiale Labiodentale	Apico-dentale	Apico-alvéolaire	Laminales	(Pré-) palatale	Vélaire (neutres)	Vélaire-lab. (Arrondie)	Labio-vélaire (Arrondies)	(neutres)	(arrondie)
occlusives	sourdes sonores	[p] [b]	[t] [d]					[k] [g]		[kp] [gb]	[ʔ]
	Occl. Pré-nas.	[mb]	[nd]		[nz]			[ŋ]	[ŋ]		
	implosives	[b]	[d]								
	affriquées				[dz]	[dʒ]					
	fricatives	[f] [v]			[s] [z]						[h]
	Rabat Labio dental	[ɸ]									
	nasales	[m]	[n]			[ɲ]		[ŋ]			
	liquides		[l]								
	latérale vibrante		[r]								
	semi-voyelles					[j]		[w]			

Source : Créée par l'auteur.

Cet inventaire phonique est quasiment identique à celui que nous observons dans les autres langues adamawa-oubanguiennes du groupe Dii ou Duru particulièrement. Les trois phones nasals peuvent tous porter la marque de syllabité. Nous pouvons voir qu'il y a cinq (05) principaux points d'articulation qui sont les bilabiales, les alvéolaires, les palatales, les vélares et la glottale. L'on note aussi la présence de 09 modes d'articulation. Les sons [nz]; [ŋ] sont obtenus dans la langue lorsque les locuteurs dugun font recours phénomène de liaison dans le parler rapide. En outre, il faut noter que sous l'influence des sons nasals /m/, /n/ et /ŋ/ en position finale d'un mot et sous l'effet de la nasalisation des voyelles ([a], [ɔ], [ɛ], [u], [ɤ], [i], [ə],) la liquide semi-voyelle [j] se prononce [ɲ]. Excepté les sons en rouge (voir tableau 9), tous les autres caractères phoniques sont phonémiques.

Le dugun compte donc 26 phonèmes consonantiques. Ces phonèmes sont représentés dans le tableau 10 ci-dessous.

**Tableau 10.**

*Tableau Phonémique des consonnes*

MODES D'ARTICULATION		POINTS D'ARTICULATION									
		Labiales		Alvéolaires			Palatales	vélares		glottales	
		Bilabiale Labiodentale	Apico-dentale	Apico-alvéolaire	Laminales	(Pré-) palatal	Vélaire (neutres)	Vélaire-lab. (Arrondies)	Labio-vélaire (Arrondies)	(neutres)	(arrondie)
occlusives	sourdes sonores	/p/, /b/	/t/, /d/					/k/, /g/		/kp/, /gb/	/ʔ/
	Occl. Pré-nas.	/mb/	/nd/								
	implosives	/b/	/d/								
	affriquées										
	fricatives	/f/, /v/			/s/, /z/						/h/
	Rabat labiodental	/ɸ/									
	nasales	/m/	/n/					/ŋ/			
	liquides		/l/ /r/								
	latérale vibrante										
	semi-voyelles					/j/		/w/			

Source : Créée par l'auteur.

#### 4.2.2 Inventaires vocaliques du dugun et leur fréquence d'emploi

La figure 14 est nous présente l'inventaire vocalique du *dugun*. Cet inventaire est proposé par l'outil dekereke. Les sons vocaliques (voir figure 14) sont au stade basique. Dekereke génère tous les sons y compris la longueur vocalique, les voyelles nasales. Le linguiste ou le chercheur averti est appelé à effectuer des analyses supplémentaires pour une étude adéquate, il doit à cet effet respecter les normes scientifiques en place.

**Figure 14.**

*Interface de tableau des inventaires des sons vocaliques(dugun)*

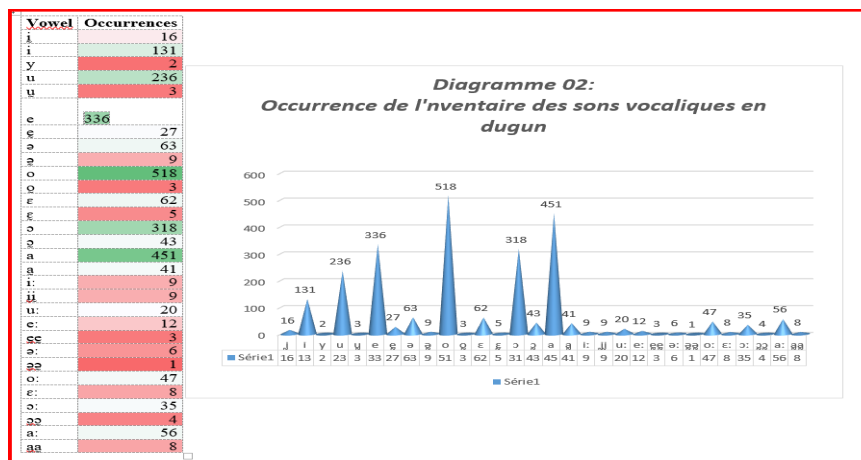
	Front Unround	Central Unround	Back Round
High [+ATR]	i		u
Mid [+ATR]	e ɛ	ə	o ɔ
Mid [-ATR]	ɛ		ɔ
Low [-ATR]		a	

*Source : Créée par l'auteur.*

La figure 15 ci-dessous quant à elle présent l'inventaire vocalique dugun selon la fréquence d'emploi et distribuée dans les mots. Cet inventaire est réalisé grâce à l'utilisation de l'outil Dekereke. Ce logiciel présente les occurrences des sons vocaliques de manière quantitative. Nous observons que le [o] vient en tête de distribution, avec 518 occurrences, il est suivi de [a] avec 451 occurrences, [e], 336 occurrences, [ɔ], 318 occurrences. etc. En suivant ainsi le tableau de ces occurrences phoniques ou phonématique, elles peuvent faciliter la tâche aux spécialistes et/ou Consultant en alphabétisation dans une perspective d'élaboration des syllabaires ou tout autre ressource didactique pour l'alphabétisation fonctionnelle.

**Figure 15.**

*Schéma de la Fréquences phonique des voyelles du dugun*



*Source : Créée par l'auteur.*

Au plan phonétique, le logiciel dekereke nous a permis d'obtenir les voyelles dugun suivantes : 8 phones dont 2 voyelles hautes, 2 voyelles mi- fermées, 3 voyelles centrales et une basse ou ouverte. Excepté [o]. Toutes ces voyelles peuvent être nasalisées. Le dugun présente des voyelles longues qui sont permanemment en opposition. Il faut préciser que cet outil organise automatiquement les ATR+ et les ATR- tout en les catégorisant dans le tableau vocalique proposé. Le tableau 11 ci-dessous est une représentation des différentes voyelles *dugun*.

**Tableau 11.**

*Voyelles phonétiques et phonémiques du Dugun*

Hautes/fermées	/i/ i		/u/ /u/
Mi- fermée	/e/ e		/o/
Centrales	/ɛ/ /ɛ/	/ə /ə/	/ɔ /ɔ/
Basses		/a/ /a/	

*Source : Créée par l'auteur.*

#### 4.2.3 Au niveau tonologique et structure syllabique

S'agissant de la structure syllabique, dekereke nous a permis de bien organiser le profile syllabique dugun selon leur type et selon la fréquence d'emploi, chose qui n'est pas pareil avec l'assistant (PA). Le tableau 12 ci-dessous est la fréquence des différentes structures syllabiques du dugun. *Dekereke* à inventorié quatorze formes syllabiques (voir tableau 12).

Au regard de la fréquence dominante, nous pouvons conclure en disant que le *dugun* est fondamentalement marqué par la structure syllabique CV, CVV, CVC, avec des formes irrégulières notamment les syllabes nasalisés(couleur orange).

**Tableau 12.**

*Structure syllabiquev*

type de syllabe	nombre total	mot-initial	mot-médiane	mots-final
CV	8090	3025	1760	3805
CVV	1960	1460	120	780
CVC	1115	955	20	965
CVVC	180	180	0	180
V	685	280	30	405
N	275	275	0	0
CCVV	10	5	5	0
CCV	85	5	50	30
VV	45	20	0	30
CNV	25	0	10	15
VC	180	175	0	160
CCCV	10	5	0	10
CNVC	5	0	0	5
CN	5	0	5	0

*Source : Créée par l'auteur.*

Au niveau tonologique, K. Snider (2020), propose une introduction extrêmement utile, claire et complète destinée aux universitaires et aux travailleurs de terrain qui souhaitent savoir comment la tonalité est exploitée dans les langues africaines et autres langues à tonalité. À priori, l'étude des tons se fonde sur des questions méthodologiques mettant en relief l'interprétation des contrastes contenus dans une langue donnée. L'auteur nous guide ensuite dans l'univers de la tonologie à travers les questions d'analyse phonologique et orthographique plus sereines. Bien que Phonology Assistant effectue des marquages des tons phonétiques, cette tâche à mon avis semble très complexe. C'est pourquoi nous suggérons que l'on utilise l'assistant (PA) et Dekereke. Dekereke ne facilite pas seulement l'étude tonale, cet outil propose des inventaires phoniques des consonnes, des voyelles, le profil syllabique avec des fréquences visibles. Surcroît, les digraphes sont également générés immédiatement. L'avantage avec l'outil dekereke est que la base de 1700 mots y est déjà intégrée, on n'a pas besoin de faire usage de FLEx. On peut faire des enregistrements audio et des transcriptions de hauteur pour mieux analyser les phénomènes tels que le rabaissement tonal, la propagation tonale, les tons flottants, etc. étudier les schèmes des racines pour identifier les tons sous-jacents, l'analyse du phénomène de l'intonation et la tonation avec une approche auto segmentale. Nous y reviendrons prochainement pour faire une étude approfondie des éléments autosegmentaux du dugun. L'outil (dekereke) donne la possibilité d'effectuer des enregistrements, sous le format « wav » (voir colonne 5, tableau 13), une méthode qui n'existe pas dans Phonology Assistant. La notation des mots enregistrés devra être renommée selon l'un des modèles suivants: 0001Corps.wav ou 0001CORPS.WAV .

Le tableau 13 ci-dessous est la description de la méthode de transcription phonétique et de segmentation des unités tonales du dugun à l'aide de Dekereke. Le tableau comporte 10 colonnes. Chaque colonne contient une information précise.

- La colonne 01(référence) représente les numéros de référence.
- La colonne 02 au fond blanc est l'ensemble de glose de mots en français (Cette colonne donne la signification des mots dans une langue d'analyse que nous appelons généralement langue source. Dans le cadre de notre analyse, les gloses sont données en français).
- La colonne 03 en gris est la transcription phonétique des mots du dugun.
- La colonne 04 en bleu (est la transcription de sons de hauteurs (tonation ou intonation au singulier).
- La colonne 05 (Fichier audio).
- La colonne 06 : Catégorie grammaticale).
- La colonne 07 en orange (est la transcription phonétique des variantes plurielles.
- La colonne 08 en vert (est la transcription de sons de hauteurs (tonation ou intonation avec les variantes plurielles.
- La colonne 9 renvoie eux schème tonal sous-jacent du radical.
- La colonne 10 le profil syllabique.

**Tableau 13.**

*Transcription phonétique et segmentation des unités tonales dugun à l'aide de Dekereke.*

Référence	Glose	Phonétique Singulier	Hauteur Singulier	Catégorie	Phon.PL	Hauteur Phon. Pluriel	Schème sous-jacent du radical	Profil Syllabique
0001	corps	[fɔn]o	[5]4	Nom	fɔn vo	[5]4	/H/	CVCV
0002	peau (d'homme)	[kɔl]ɔ	[5]4	Nom	kɔl vo	[5]4	/H/	CVCV
0003	tête	[jul]u	[5]4	Nom	jul vo	[5]4	/H/	CVCV
0004	front	[gbɛ:g]o	[C ]4	Nom	gbɛ:g vo	[- ]4	/H/	CVVCV
0005	visage	[gbɛ:tun]o	[35]4	Nom	gbɛ:tun vo	[35] 3	/H/	CVV CVCV
0006	œil	[nɔʔ]ɔ	[6 5]	Nom	nɔʔɔ	[5]4	/H/	CVCV
0007	sourcil	[nɔʔgagar]o	[53 3]4	Nom	nɔʔ vo	[5]3	/H/	CVCCVC VCV
0008	paupière	[nɔʔ lɛ:r bɛr]o	[5 4 3]4	Nom	nɔʔ lɛ:vo	[5]1	/BH/	CVCCVV CCVCV

*Source : Créée par l'auteur.*

#### 4.2.4 Au niveau des utilisateurs (enseignant, étudiants, chercheurs, etc.)

Pour une utilisation efficiente des outils du TAL, nous proposons l'intégration des technologies linguistiques dans les Universités au Cameroun, la formation axée sur les compétences des enseignants et étudiants par l'entremise d'un parcours académique et professionnel en ingénierie linguistique. De nos jours, l'on parle des humanités numériques. Pour Mounier (2018), les humanités numériques sont un ensemble de pratiques méthodologiques et de transformations dans le domaine des sciences humaines et sociales à l'ère du numérique. Par ailleurs, la réussite d'implémentation des outils du TAL nécessite une mise en œuvre d'un mécanisme collaboratif sérieux : un comité consultatif constitué des experts, des politiques d'enseignement sérieux, l'implication des partenaires spécialisés, des disciplines connexes, un laboratoire opérationnel et équipé.

#### Conclusion

En définitive, l'expérimentation de Phonology Assistant, appliqué au dugun au Cameroun, a abouti aux résultats majeurs suivants : la méthode de description et d'analyse automatique, avec la nécessité de combiner plusieurs outils de TAL pour des analyses satisfaisantes, mais aussi de vulgariser ces outils dans les universités au Cameroun. Le premier résultat est centré sur des méthodes de transcription

phonétique, analyse phonologique notamment, la fluidité et la cohérence dans la segmentation des unités tonales et autosegmentales, l'élaboration systématique des inventaires phoniques, des paires minimales, l'environnement identique, le profile syllabique. La description régulière des variations des allophones, la documentation et la sauvegarde des valeurs linguistiques des langues à faibles ressources comme le dugun. Le deuxième résultat est axé sur l'interopérabilité des outils TAL suite aux insuffisances constatées sur l'utilisation de l'assistant (PA): D'une part l'inexistence et la complexité de certaines fonctionnalités, l'interface essentiellement en anglais ou espagnol. D'autre part l'ignorance et l'utilisation inappropriée de cette technologie linguistique par les utilisateurs. Au regard des limites relevées, il a été suggéré une utilisation complémentaire des outils comme FLEx et Dekereke associés à l'assistant (PA) et l'implémentation de ces outils dans les Universités au Cameroun. Ce travail contribue à la standardisation d'une langue peu documentée et à l'enseignement des langues à travers des outils du TALN. Cet intérêt pourrait justifier la pertinence de la linguistique du développement telle que perçue par Métangmo (2019), dans le processus d'amélioration des conditions didactiques et pédagogiques des étudiants ?

---

**Niveau(x) de Contribution des Auteur(s):** A.B. 100 %

**Approbation du Comité d'Ethique:** Il ne s'agit pas d'une étude qui nécessite l'approbation du comité d'éthique.

**Soutien Financier:** Aucun soutien financier n'a été reçu pour l'étude.

**Conflit d'Intérêts:** Il n'y a aucun conflit d'intérêts dans l'étude.

**Author Contributions:** A. B. 100 %

**Ethics Committee Approval:** This study does not require ethics committee approval.

**Financial Support:** The study did not receive financial support.

**Conflict of Interest:** There is no conflict of interest in the study.

---

## Bibliographie

Bird, S. (2009). Natural Language Processing and Linguistic Fieldwork. *Computational Linguistics*. 35(3), 469–474. <https://doi.org/10.1162/coli.35.3.469>

Dingemans, M. (2008). Phonology Assistant 3.0.1 from SIL International. *Phonology Assistant Review*. 2(2), 325-331. <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/52cff471-0b9d-47d2-99bc-c1fcb5cb1a1a/content>

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

Essono, J.M. (2006). *Phonétique, phonologie et morphophonologie*. Cameroon University Press.

- Kempton, T., & Moore, R. K. (2009). Finding allophones: An evaluation on consonants in the TIMIT corpus. In *Proceedings of Interspeech*, 1651-1654. [https://www.isca-archive.org/interspeech\\_2009/kempton09\\_interspeech.pdf](https://www.isca-archive.org/interspeech_2009/kempton09_interspeech.pdf)
- Lebart, L., & Salem, A. (1994). *Statistique textuelle* (2e éd.). Dunod.
- Lode, L. (1992). *Aperçu des langues parlées dans le département du Faro*. Document inédit, Cameroun.
- Mboning, E. & Wandji, O. (2021). Construire des ressources collaboratives pour les langues peu dotées : Une modélisation orientée communauté. In P. Denis, N. Grabar, A. Fraisse, R. Cardon, B. Jacquemin, E. Kergosien, A. Balvet (Eds) *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1: conférence principale* (pp. 125-133), ATALA.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Métangmo, T. L. (2019). *Pour une linguistique du développement : Essai d'épistémologie sur l'émergence d'un nouveau paradigme en sciences du langage*. Édition Sciences et Bien Commun.
- Mohamed, Z. K. (2016). *Natural Language Processing and Computational Linguistics I Speech, Morphology and Syntax*. ISTE.
- Mounier, P. (2018). *Les humanités numériques: Une histoire critique*. Maison des Sciences de l'Homme.
- SIL International (2019). Language Technology for Resource-Scarce Languages. <https://software.sil.org/>
- SIL International. (2012). FieldWorks Language Explorer (FLEx) & Toolbox User Documentation. <https://software.sil.org/fieldworks/>
- SIL International. (2023). *Phonology Assistant. Overview and Tutorials*. <https://software.sil.org/phonologyassistant/>
- Snider, K., & Roberts, J. (2004). *SIL comparative african wordlist (SILCAWL)*. SIL International. [https://www.eva.mpg.de/lingua/tools-at-lingboard/pdf/Snider\\_silewp2006-005.pdf](https://www.eva.mpg.de/lingua/tools-at-lingboard/pdf/Snider_silewp2006-005.pdf)
- Snider, Keith. (2020). *Tone Analysis for Field Linguists*. (2e éd.). SIL International Publications.
- Teronpi, M. (2025). A Preliminary Description of the Phonology of Ri-Bhoi Amri Karbi. *International Journal of Humanities and Social Science Invention*, 14(1), 40-47. <https://doi.org/10.35629/7722-14014047>.