Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles
Lille, France, 28 juin au 2 juillet 2021

Construire des ressources collaboratives pour les langues peu dotées: une modélisation orientée communauté

Elvis Mboning Tchiaze[1, 2] Ornella Wandji[1]
(1) NTeALan Research and Development, Makepe, Douala, Cameroun
(2) ERTIM, 2 Rue de Lille, Paris, France
levismboning@ntealan.org, ornella.wandji@ntealan.org

RÉSUMÉ

Les applications du traitement automatique des langues (TAL) nourrissent aujourd'hui une bonne partie des langues indo-européennes en raison des corpus linguistiques de qualité disponibles en grande quantité et variété. Les corpus de données open sources en langues africaines étant quasi inexistants, comment arrimer les avancées du TAL à ces langues peu dotées ? Dans cet article, nous examinons le problème de construction des ressources lexicographiques pour les langues peu dotées. Nous souhaitons introduire un modèle de construction des ressources lexicographiques en exploitant les compétences socio-linguistiques des communautés linguistiques locales. Au fil des sections, nous présenterons le nouveau modèle de codification des dictionnaires issue de cette modélisation orientée communauté.

ABSTRACT

Building collaborative resources for poorly endowed languages : community-oriented modeling

The applications of natural language processing (NLP) today feed a large part of Indo-European languages, with a large body of quality data available in large quantities. As open source, data corpora in African languages are almost non-existent, how can the advances in NLP be secured for these poorly endowed languages ? In this article, we address the problem of constructing lexicographic resources. We wish to introduce a model for building lexical resources by exploiting the socio-linguistic skills of local linguistic communities. Throughout the sections, we will present the new dictionary coding model resulting from this community-oriented modeling.

MOTS-CLÉS : Langues africaines, lexicographie électronique, NTeALan, modèle collaboratif, graphe, modèle basé sur la communauté.

KEYWORDS: African languages, electronic lexicography, NTeALan, collaboration model, graph, community-based model.

1 Introduction

Il y a encore quelques années de cela, les travaux de développement de ressources en langues africaines ne faisaient pas l'objet d'affluence dans le monde de la recherche. Mais de nos jours, des chercheurs, équipes de recherche, laboratoires, universités autant en Afrique qu'en Occident, et parfois en collaboration, se consacrent de plus en plus à la numérisation des langues africaines, à la conception de dictionnaires électroniques et d'autres ressources et outils du TAL dans et pour

Volume 1 : conférence principale, pages 125–133.

ces langues. L'association NTeALan Social Network [1], à travers sa jeune équipe de recherche, veut s'inscrire dans cette lignée. Depuis 2018, elle développe un modèle open source de construction collaborative des ressources lexicographiques pour les langues africaines.

En exploitant le modèle collaboratif de (Holtzblatt & Beyer, 2017), nous souhaitons dans cet article, présenter les premiers résultats des travaux de refonte des données lexicographiques de nos plateformes collaboratives construites sous un modèle orienté communauté. Il s'agit précisément de la description du nouveau modèle de codification de nos dictionnaires et l'encapsulation du modèle collaboratif vers un modèle orienté communauté inspiré des graphes. Dans les sections ci-dessous, nous présenterons d'abord les propriétés de l'ancien format de codification, ensuite celles du nouveau format et enfin la nouvelle modélisation orientée communauté.

2 XND : format initial de NTeALan

Le format XND (XML NTeALan Dictionaries) est un format de codification intermédiaire ayant une porté morpho-syntaxique, développé à partir de 2018 pour codifier et outiller les dictionnaires bilingues produits par les plateformes collaboratives de NTeALan (Mboning Tchiaze *et al.*, 2020; Mboning *et al.*, 2020).

2.1 Pourquoi créer un nouveau format de codification ?

Le choix de créer un format pour codifier nos données lexicographiques s'est fait après plusieurs observations et essais techniques exploités dans la littérature (Bosch & Pretorius, 2002, 2003; Heid, 2014; Pretorius & Bosch, 2003; Kotzé, 2005a,b; Bosch & Pretorius, 2004; Prinsloo, 2012; Bosch & Pretorius, 2011; Pretorius & Bosch, 2012; Nogwina *et al.*, 2013; Prinsloo *et al.*, 2012; Benoit & Turcan, 2006).

En effet, chaque plateforme de gestion de ressources lexicales possède son propre modèle de structuration et de présentation des données, c'est le cas des plateformes suivantes : `Kosh` (Mondaca *et al.*, 2019), `ELEXIS Dictionary Service` et `Djibiki` (Mangeot, 2006). Parmi les formats utilisés [2] pour codifier ces données, le format XML (principalement les normes TEI, CES et LMF) est aujourd'hui un choix de référence pour la structuration de données linguistiques, lexicographiques et terminographiques. Malheureusement, ces normes ne sont pas souvent adaptées pour représenter et décrire certaines particularités morpho-syntaxiques des langues africaines [3]. Pour preuve, plusieurs phénomènes linguistiques, tels que le concept de classe nominale, la notion de clics ou encore la gestion de la traduction et de la localisation des variantes dialectales de l'entrée d'article (mot vedette) ne sont pas traités explicitement, malgré tous les besoins exprimés en la matière [4].

2.2 Description du format XND : version initiale

En analysant la structure d'une langue sémi-Bantu (`yemba` parlée dans la région de l'Ouest au Cameroun), nous avons initialement décidé de définir un modèle propriétaire de structuration XML, un intermédiaire entre l'environnement interne de NTeALan et les normes externes, dont la structure

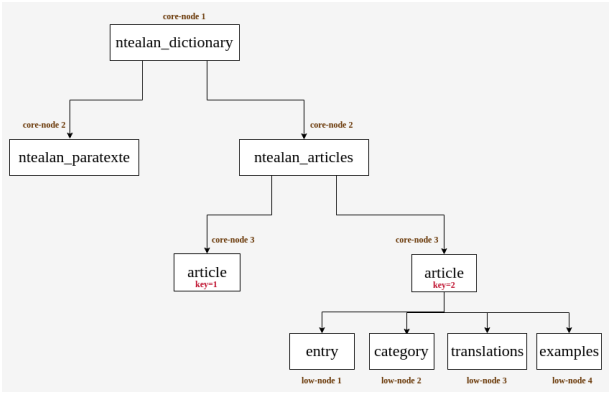
1. Elle a été initialement légalisée en 2019 au Cameroun sous le nom NTeALan. Site web officiel : https://ntealan.org
2. On peut citer les formats TEI Lex-0 (Romary & Tasovac, 2018) et Lexicog (OntoLex Lemon Lexicography du W3C), plus récents, qui sont fréquemment utilisés pour les codifier.
3. Certainement pour les mêmes raisons, aucuns des auteurs des travaux précédents sur l'xmlisation (mise au format XML) des langues africaines, n'a essayé ces formats.
4. Néanmoins, il faut noter qu'il est possible dans ces standards d'ajouter de nouvelles classes (balises et attributs) en complément des classes existantes.

s'inspirerait des 4 grandes familles de langues africaines, à savoir : la famille Afro-asiatique, la famille Niger-Kordofaniene (anciennement appelée Niger-Congo), la famille Nilo-Saharienne et la famille Khoisane. Ainsi, lors de la description de ce format, trois principes ont guidé nos choix : `représentation` (description en composants linguistiques), `simplification` (arbre et noms des balises explicites) et `extensibilité` (ouverture aux nouveaux noeuds). Cf. la figure (a) du tableau 1 :

— **Représentation** : avec ce principe, nous décrivons les données du langage au plus petit niveau morpho-syntaxique bantu, c'est-à-dire les composants du mot (variante dialectale = `préfixe + radical + suffixe`) et les composants de la phrase comme l'accord de classe (1/2, 3/4, 5/7, etc.), les préfixes d'accord. À noter que c'est l'accord de classe qui régit les structures syntaxiques dans les langues bantu et semi-bantu au sens triste du terme.
— **Simplification** : nous choisissons ici des noms de balises XML dans une langue nationale locale facilement compréhensibles par la communauté des utilisateurs. De plus, nous avons choisi d'utiliser une représentation XML linéaire, avec moins d'ascendants/descendants pour privilégier plus d'enfants du même nœud parent.
— **Extensibilité** : nous donnons aux contributeurs externes la possibilité d'étendre nos principales structures XML en ajoutant de nouveaux nœuds (enfants ou nœuds parents), en fonction de l'élément à représenter.

Plus concrètement, le nœud (balise XML) racine (*core-node*) `<ntealan_dictionary>`, est divisé en deux sous-nœuds : `<ntealan_paratexte>` et `<ntealan_articles>`. La balise `<ntealan_paratexte>` décrit les métadonnées autour de la ou des version(s) du dictionnaire (contexte de production du dictionnaire, information sur la source des données, les auteurs, l'année de création, les droits d'auteurs, etc.). Et la balise `<ntealan_articles>` décrit tous les articles du dictionnaire avec la balise enfant (`<article>`). Cf. la figure (b) du tableau 1.

(a)



(b)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ntealan_dictionary SYSTEM "DTD_NTeALan_dictionaries.dtd">

<ntealan_dictionary>

  <ntealan_paratexte> ... </ntealan_paratexte>

  <ntealan_articles> ... </ntealan_articles>

  <ntealan_tableaux_noms>
    <tabl_NP> ... </tabl_NP>
    <tabl_toponymes> ... </tabl_toponymes>
  </ntealan_tableaux_noms>

</ntealan_dictionary>
```

TABLE 1 – (a) Premier modèle de représentation initial des dictionnaires de NTeALan. (b) Première structure formelle du format XND de NTeALan

Chaque noeud article `<article>` a ses propres sous-nœuds : `<entry>` (entrée ou mot vedette de l'article constituée d'une ou plusieurs variantes dialectales `<variant>`, elle aussi est constituée des noeuds `<prefix> | <radical> | <suffix>`, partie intégrante du mot vedette), `<category>` (catégorie grammaticale en lien avec les variantes dialectales), `<classe_d_accords>` (classe d'accord `<cl_sing>` / `<cl_plur>` d'une entrée d'article de type Nom ou dans certains cas de type Adjectif), `<conjugaison>` (forme conjuguée d'une entrée d'article de type Verbe), `<translations>` (traductions associées aux variantes dialectales), `<examples>` (contextualisation des variantes dialectales).

Pour les dictionnaires ayant des entités nommées (noms de personnes et de lieux), une structuration minimale a été proposée à travers le noeud `<ntealan_tableaux_noms>`.

Il faut remarquer que le format XND, bien que spécifique aux plateformes de NTeALan, est exportable vers quelques standards existants : à ce jour il s'agit du format TEI P5 et du format LMF. Cette organisation nous permet de passer d'un format à l'autre sans perte d'informations, avec peu être une mutation structurelle et ce en fonction des outils TAL manipulés.

2.3 Limites du format XND

Après plus de deux années d'existence, les données sur la plateforme sont passées de 12 000 entrées d'article de dictionnaire à plus de 34 000 entrées. Pour être plus précis, 24 dictionnaires ont été crées avec un nombre d'articles compris entre 0 et 12 000. Les données sont accessibles via une API REST open source [5] et deux plateformes web [6].

La simplicité et la linéarité du format XND telle que voulue dès le départ a servi avec exemplarité les premières vagues d'xmlisation de dictionnaires créées par les communautés de locuteurs locaux et étudiantes. Cependant, à son état actuel, ce format ne permet pas de se projeter sur le modèle collaboratif de création des ressources synchrones ou asynchrones. Plusieurs raisons peuvent expliquer ce constat :

— Les propriétés linguistiques des 4 grandes familles des langues africaines sont sous-représentées dans le format actuel.
— La gestion des communautés de contributeurs n'a pas été codifiée dans ce format, bien qu'elle soit prise en compte dans la plateforme des dictionnaires.
— La gestion des applications externes et la standardisation ont été au coeur des enjeux de la définition du format XND.
— Le moteur de recherche appliqué à ce format n'est pas si efficace dans la recherche d'informations voulues par l'utilisateur.

Au vu de ceci, il était plus que nécessaire de faire évoluer le format XND pour mieux servir les langues africaines et leurs utilisateurs. Il devrait aussi continuer à respecter les 3 principes de base (cf. section 2.2). qui ont sous-tendu sa création.

3 Modélisation d'un nouveau format de codification des dictionnaires de NTeALan

En Afrique, d'après (Assoumou, 2017), «l'individu n'existe que pour sa communauté, cette dernière est au service de tous et de chacun dans un environnement où la solidarité, la fraternité et le respect sont des maîtres mots. Des coutumes et les traditions fondent le mode de vie des populations. Elles constituent un code social, juridique, moral, ..., hérités de la sagesse ancestrale et dont les lois font des communautés des États de droit.»

3.1 Vers un modèle collaboratif orienté communauté

Le concept de communauté n'est pas un choix anodin dans notre cas. En effet, la sociologie africaine est construite sur le modèle de la communauté, c'est-à-dire un ensemble de groupes sociaux et de

5. https://apis.ntealan.net/ntealan/dictionaries
6. https://ntealan.net/dictionaries-platform, https://ntealan.net/dictionaries

sous-groupes partageant une même langue, une même culture et un même espace géographique. Dans ces groupes, la solidarité se crée et des actions sociales émergent pour l'intérêt de tous. Ce concept montre clairement le lien culturel fort qui unit chaque citoyen à sa communauté, avant même celle de son pays (Tunde, 2012).

Notre nouveau modèle collaboratif puisera ses forces dans les sociétés communautaires africaines. Autrement dit, il exploitera les compétences socio-linguistiques locales propre à chaque communauté pour construire des ressources lexicographiques inclusives. Nous nous inspirons de l'organisation des comités de langues [7] au Cameroun pour structurer ces communautés. À chaque communauté, nous associons : les experts linguistes du comité, les enseignants/chercheurs, les locuteurs natifs et la diaspora d'une langue. Ces communautés travaillent ensemble pour créer des ressources dans leur langue au sein d'un espace collaboratif en ligne.

Cet espace collaboratif (espace de partage de compétences socio-linguistiques locales) doit répondre à la fois à toutes les exigences de ce modèle communautaire, aux exigences techniques (ergonomie, codification des données, RGPD, etc) et aux exigences scientifiques des domaines exploités (linguistique africaine, lexicographie électronique, linguistique de corpus, pédagogie/didactique, TAL). Cf. l'illustration 1.

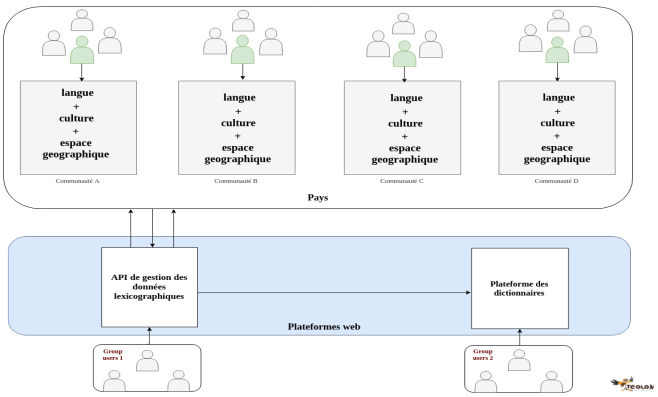


FIGURE 1 – Modèle collaboratif orienté communauté : de la constitution sociale des communautés à leur implication dans les plateformes collaboratives en ligne.

3.2 Encapsulation dans le nouveau format de codification de NTeALan

Pour ce dernier point, la représentation des données par les graphes devient une pratique assez récurrente aujourd'hui pour modéliser les connaissances ayant des liens structuraux et fonctionnels identifiables. Comme on l'a vu dans les sections précédentes, plusieurs formalismes existent sur le marché technologique, chacun ayant ses particularités, ses avantages et ses inconvénients. Comment combiner ou faire évoluer un modèle de représentation natif (le format XND initial), en un modèle de codification sémantique ouvert aux principes de la collaboration et du partage, tout en respectant le cadre socio-culturel des locuteurs natifs des langues concernées ?

Avant d'aller plus loin dans cette modélisation, nous avons d'abord voulu proposer une version améliorée de notre format XND. Il fallait donc répondre à quatre nouveaux objectifs :

7. Pour prendre l'exemple du Cameroun, chaque langue nationale a un comité de langue qui se charge de sa standardisation et de son développement. Tous ces comités de langues sont réunis autour d'une association nationale dénommée ANACLAC (Association Nationale des Comités de Langues Camerounaises).

— Normaliser l'architecture lexicographique avec l'introduction de nouveaux composants linguistiques issus de la P5 de la TEI, des entités nommées, etc.
— Standardiser les noms de balises avec un nommage en *anglais*
— Trouver les liens explicites (connexions lexicales, sémantiques, structurelles et syntaxiques) entre les balises et leurs attributs afin de créer des sous-réseaux à l'intérieur de tous les composants de l'article à la famille de langue en passant par le dictionnaire.
— Créer un système de gestion de version greffée à chaque composant afin de suivre les modifications des contributeurs en mode collaboratif.

La figure 2 montre la nouvelle représentation des composants d'un dictionnaire codifié avec le format XND. Pour mieux comprendre ce tableau, nous allons procéder par ses propriétés :

— **Les couleurs** : Les éléments en `fond jaune` représentent les balises modifiées (cf. `plur_cl, ?class_accord, source_year, +plur_cl, source_links, ?ambig_cl, +conj_form, purpose, ?plur_group, build_context, authors_version, ntealan_paratext`, etc). Les éléments en `fond vert` représentent les balises ajoutées (cf. `+date, title, ?place_table, ?definition, dictionary_name, nb`, etc.) et les éléments en `fond gris` représentent les balises existantes non modifiées (cf. `source, entry, variant, examples, institution, +article, +equivalent`, etc). Les textes écrits en `vert sous un fond blanc` représentent les attributs de balise. Entre crochet, le nom de l'attribut suivit du type numérique de sa valeur `fond gris` (cf. `[ref]:string, [etym]:string, [usg]:string`, etc) ou `string` équivaut à chaîne de caractères.
— **Les blocs** : les blocs avec une entête de couleur `noire foncée` représentent les noeuds principaux d'un point de vue structurel (cf. `articles, examples, conjugaison, ntealan_entities, entry`, etc.) et ceux avec des entêtes de couleur `rouge foncée` représentent les sous-noeuds. Certains blocs de couleurs rouges peuvent décrire le type de sa valeur (cf. `+author, +file_link`, etc), les attributs de la balise référencée (cf. `+noun, +place`) ou une sous-référence (cf. `+variant`).
— **Les symboles** : ont la même signification que les quantificateurs des expressions régulières (*****=zéro ou plusieurs fois, **+**=une ou plusieurs fois et **?**=zéro ou une fois)
— **Les connexions** : si d'une part les connexions non fléchées permettent d'établir un lien de parenté (parent / enfant) et de références (parent / ref) entre les noeuds, les connexions fléchées d'autre part permettent de décrire le contenu de l'élément fléché à partir de son parent.

Nous appelons `noeuds` tous les blocs du schéma de la figure 2 qui portent une information lexicographique définie. Les noeuds de cette nouvelle version du format XND sont organisés en deux types : le type `tag` équivaut à une balise ou un bloc (cf. `?noun_table:tag, translations:tag`, etc.) et le type `string` équivaut au contenu de la balise (cf. `author_id:string, dictionary_name:string`, etc.). Le type `string` a été généralisé sur toutes les balises pour normaliser leur contenu. Un transcodage (casting en anglais) permettrait de revenir au type initial (Exemple : passer du type 'string' `[nb]:"1"` au type 'integer' `[nb]:1`).

Un point d'honneur a été mis sur la balise du mot vedette de l'article décrivant les variantes dialectales associées [8] (`variant`). En nous appuyant sur les propriétés de la `XML TEI P5` et des travaux de (Bosch *et al.*, 2007; Bosch & Pretorius, 2011; Khoule *et al.*, 2016) nous l'avons enrichi de plusieurs

8. Nous avons choisi rester sur une description des variantes pour chaque mot vedette parce qu'elles nous permettent de mieux observer les différences linguistiques entre les sous-communautés de cette langue. Ces propriétés linguistiques seront d'un grand apport pour les taches de désambiguisation lexicale généralement constaté dans la construction des outils TAL.

nouvelles informations linguistiques afin de mieux cerner sa compréhension linguistique. Entre autres nouvelles propriétés, nous avons le `[form]` (information morphologique), `[sem]` (information sémantique), `[pron]` (prononciation), `[usg]` (cas d'usage discursif du mot), `[case]` (information sur le cas grammatical), `[syll]` (forme syllabique du mot), `[gen]` (le genre si le mot est un nom).

Il en est de même pour les noeuds `equivalent`, `example` avec les nouveaux attributs (`[syn]`, `[author]`, `[nb]`), le noeud `conj_form` avec `[mood]`, `[per]`), le noeud `cat` avec `[gen]`, `[subc]`).

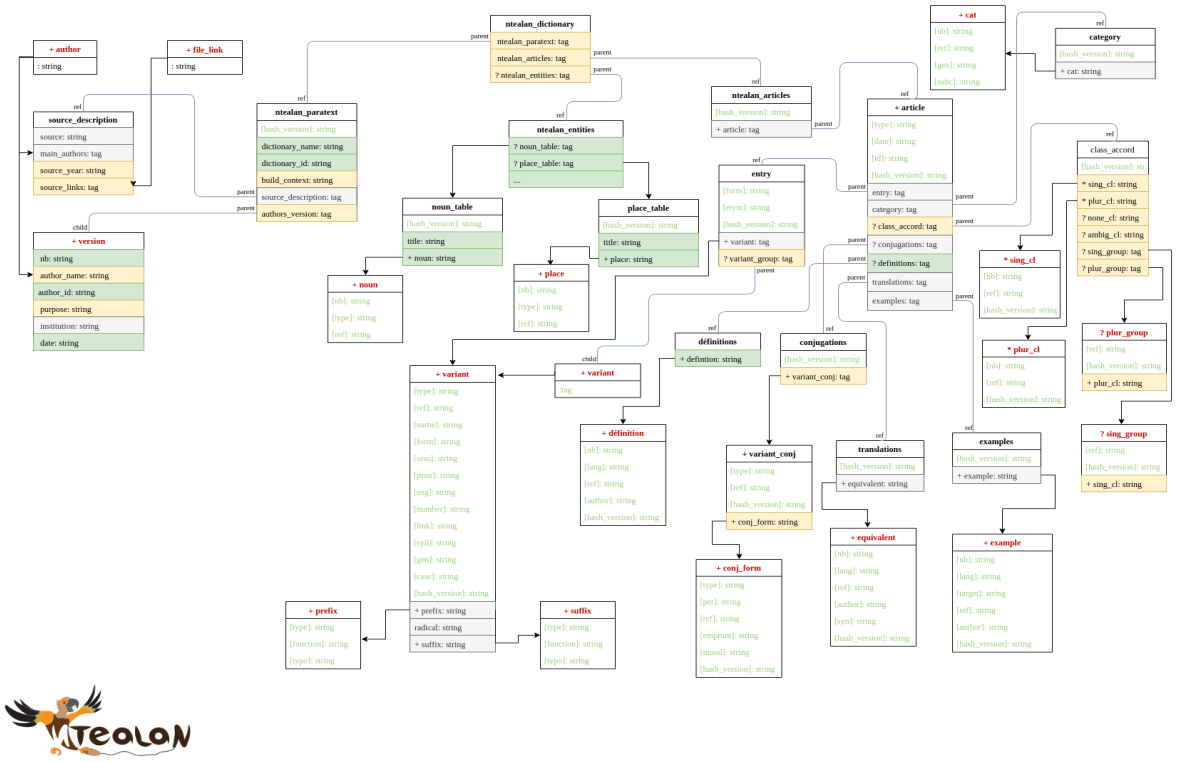


FIGURE 2 – Représentation de l'architecture améliorée du format XND et les liens entre les composants d'un dictionnaire. Regroupé en famille de langues, les dictionnaires forment un réseau.

4 Conclusion

Dans cet article, nous avons introduit un nouveau modèle collaboratif en ligne basé sur les communautés. Ce modèle orienté communauté nous permet d'associer les communautés de langues, les locuteurs natifs, chercheurs et la diaspora au sein d'un même espace de création et de partage de ressources linguistiques. Avec cette réorganisation, la nouvelle codification permettra de mieux servir les données construites par les communautés et on pourra alors envisager s'ouvrir à d'autres types de ressources autres que lexicographiques. Une mise à jour prochaine de nos plateformes permettra de déployer ce nouveau modèle afin de l'évaluer auprès de nos membres et des communautés à constituer.

Remerciements

Ce travail a été rendu possible grâce au support financier de l'équipe ERTIM de l'INALCO dans le cadre d'un projet de recherche portant sur la refonte globale du modèle collaboratif et de codification des ressources lexicographiques gérées par les plateformes collaboratives de NTeALan.

Références

ASSOUMOU J. (2017). *Culture et développement en Afrique : des perles et des pourceaux ?*, In J. ASSOUMOU & F. AMABIAMINA, Éds., *Pour une culture africaine au service du développement. Des industrie culturelles viables pour une croissance durable*, p. 14–34.

BENOIT J.-L. & TURCAN I. (2006). La TEI au service de la transmission documentaire ou de la valorisation des richesses patrimoniales : le cas difficile des dictionnaires anciens. *ANAGRAM' 2006 : Atelier sur la numérisation de l'Ecrit Ancien et des GRAndes Masses de données*.

BOSCH S. & PRETORIUS L. (2002). Finite-state computational morphology-treatment of the zulu noun. *South African computer journal*, **2002**(28), 30–38.

BOSCH S. & PRETORIUS L. (2011). Towards zulu corpus clean-up, lexicon development and corpus annotation by means of computational morphological analysis. *South African Journal of African Languages*, **31**(1), 138–158.

BOSCH S. E. & PRETORIUS L. (2003). Towards technologically enabling the indigenous languages of south africa : the central role of computational morphology. *Interactions of the Association for Computing Machinery 10 (2) (Special Issue : HCI in the developing world)*, p. 56–63.

BOSCH S. E. & PRETORIUS L. (2004). Software tools for morphological tagging of zulu corpora and lexicon development. In *LREC*.

BOSCH S. E., PRETORIUS L. & JONES J. (2007). Towards machine-readable lexicons for south african bantu languages. *Nordic Journal of African Studies*, **16**(2).

HEID U. (2014). Natural language processing techniques for improved user-friendliness of electronic dictionaries. In *of the XVI Euralex International Congress : The User in Focus*, p. 47–62.

HOLTZBLATT K. & BEYER H. (2017). *7 - Building Experience Models*. Interactive Technologies. Boston : Morgan Kaufmann. DOI : 10.1016/B978-0-12-800894-2.00007-7.

KHOULE M., MANGEOT M., NGUER E. H. M. & CISSÉ M.-T. (2016). iBaatukaay : un projet de base lexicale multilingue contributive sur le web à structure pivot pour les langues africaines notamment sénégalaises. In *Atelier Traitement Automatique des Langues Africaines TALAf 2016, conférence JEP-TALN-RECITAL 2016*, Paris, France.

KOTZÉ A. E. (2005a). Towards a morphological analyser for past tense forms in northern sotho : verb stems with final'm'and'n'. *Southern African linguistics and applied language studies*, **23**(3), 245–258.

KOTZÉ P. M. (2005b). A finite-state transducer for northern sotho deverbative nouns : the morpho-phonemic rules. *Southern African linguistics and applied language studies*, **23**(4), 393–403.

MANGEOT M. (2006). Dictionary building with the jibiki platform. In C. O. ELISA CORINO, CARLA MARELLO, Éd., *Proceedings of the 12th EURALEX International Congress*, p. 185–188, Torino, Italy : Edizioni dell'Orso.

MBONING E., BALEBA D., BASSAHAK J. M., WANDJI O. & ASSOUMOU J. (2020). NTeALan Dictionaries Platforms : An Example Of Collaboration-Based Model. In *Proceedings of the 1st International Workshop on Language Technology Platforms*, p. 66–72, Marseille, France : European Language Resources Association.

MBONING TCHIAZE E., BASSAHAK J. M., BALEBA D., WANDJI O. & ASSOUMOU J. (2020). Building Collaboration-based Resources in Endowed African Languages : Case of NTeALan Dictionaries Platform. In *Proceedings of the first workshop on Resources for African Indigenous Languages*, p. 51–56, Marseille, France : European Language Resources Association (ELRA).

MONDACA F., SCHILDKAMP P. & RAU F. (2019). Kosh : Apis for lexical data. https://github.com/cceh/kosh.

NOGWINA M., SHIBESHI Z. & MALI Z. (2013). Towards developing a stemmer for the isixhosa. *WIP. SATNAC Conference*.

PRETORIUS L. & BOSCH S. (2012). Semi-automated extraction of morphological grammars for nguni with special reference to southern ndebele. *Language Technology for Normalisation of Less-Resourced Languages*, p.73̃.

PRETORIUS L. & BOSCH S. E. (2003). Finite-state computational morphology : An analyzer prototype for zulu. *Machine Translation*, **18**(3), 195–216.

PRINSLOO D. (2012). Lexicography in non-european languages. *The Encyclopedia of Applied Linguistics*.

PRINSLOO D. J., HEID U., BOTHMA T. & FAASS G. (2012). Devices for information presentation in electronic dictionaries. *Lexikos*, **22**, 290–320.

ROMARY L. & TASOVAC T. (2018). TEI Lex-0 : A Target Format for TEI-Encoded Dictionaries and Lexical Resources. In *TEI Conference and Members' Meeting*, Tokyo, Japan.

TUNDE O. (2012). Investigating the Language Situation in Africa. In *Language and Law*, Language rights, p. 272–293. Great Clarendon street : Oxford Handbooks in Linguistics.