Proceedings of the First workshop on Resources for African Indigenous Languages (RAIL), pages 51–56
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Building Collaboration-based Resources In Endowed African Languages: Case Of NTeALan Dictionaries Platform

Elvis MBONING[1 2], **Daniel BALEBA**[1], **Jean Marc BASSAHAK**[1],**Ornella WANDJI**[1]
NTeALan[1], ERTIM (INALCO)[2]
Tradex Makepe - Douala (Cameroon), 2 rue de Lille - Paris (France)
elvis.mboning@inalco.fr[2]
{levismboning, daniel.baleba, bassahak, ornella.wandji}@ntealan.org[1]

Abstract

In a context where open-source NLP resources and tools in African languages are scarce and dispersed, it is difficult for researchers to truly fit African languages into current algorithms of artificial intelligence. Created in 2017, with the aim of building communities of voluntary contributors around African native and/or national languages, cultures, NLP technologies and artificial intelligence, the NTeALan association has set up a series of web collaborative platforms intended to allow the aforementioned communities to create and administer their own lexicographic resources. In this article, we present on the one hand the first versions of the three platforms: the REST API for saving lexicographical resources, the dictionary management platform and the collaborative dictionary platform; on the other hand, we describe the data format chosen and used to encapsulate our resources. After experimenting with a few dictionaries and some users feedback, we are convinced that only collaboration-based approach and platforms can effectively respond to the production of good resources in African native and/or national languages.

Keywords: African languages, NLP, resources, xmlisation, collaboration, dictionaries, lexicography, open-source

1. Introduction

Language plays an important role in defining the identity and humanity of individuals. As Tunde Opeibi (Tunde, 2012) said "In Africa, evidence shows that language has become a very strong factor for ethno national identity, with the ethnic loyalty overriding the national interest". To date, the African continent has more than 2000 languages, more than two thirds of which are poorly endowed. Among the reasons justifying this observation, we can list:

- The lack of a strong linguistic policy in favor of these languages
- The absence of the majority of these languages in the digital space (social networks, online or mobile platform, etc.) and in the educational system (mainly described by (Tadadjeu, 2004) and (Don, 2010))
- The lack of open-source African linguistic resources (textual and oral), Natural Language Processing (NLP) and/or Natural Language Understanding (NLU) tools available for most of these languages
- The lack of experts in NLP, NLU and Artificial Intelligence (AI) trained in the continent and who are specialists in these languages
- The lack of open-source African linguistic resources (textual and oral) and NLP and/or NLU tools available for most of these languages

For several years now, artificial intelligence technologies, including those of NLP, have greatly contributed to the economic and scientific emergence of poorly endowed languages in northern countries, thanks to the availability of lexicography and terminography resources in sufficient quantity. African languages benefit very little from these intelligent tools because of the scarcity of structured data and collaborative platforms available for building linguistic and cultural knowledge bases. In order to meet this need and complement the initiatives already present on the continent ((De Pauw et al., 2009), (Mboning, 2016), (Vydrin, Valentin and Rovenchak, Andrij and Maslinsky, Kirill, 2016), (Abate et al., 2018), (Mboning, Elvis and NTeALan contributors, 2017), (Mangeot and Enguehard, 2011), (De Schryver, 2010), Afrilex association (Ruthven, 2005)), and also those from African, European and American research centers, NTeALan (New Technologies for African Languages), specialized in the development of NLP and NLU tools for teaching African languages and cultures, has set up a collaborative and open-source platform for building lexical resources for African national languages. Our main goal is to deal with languages spoken in French-speaking African countries.

This paper focuses on the development of African linguistics and cultural resources, which is an important starting point for the technological step forward of each African language. We describe our collaborative language resources platform focusing on lexicographic data. This platform is divided into three components: the open-source dictionary backup API (back-end), the dictionary management platform and the collaborative dictionary platform (fronts-end).

2. Context of the work

2.1. NTeALan project

Created in 2017[1] and managed by academics and the African Learned Society, NTeALan is an Association that

[1] Namely by Elvis Mboning (NLP Research Engineer at INALCO) and Jean Marc Bassahak (Contractor, Web designer and developer), who were later on joined by Jules Assoumou, Head of

works for the implementation of intelligent technological tools, for the development, promotion and teaching of African native and/or national languages. Our goals are to digitize, safeguard and promote these poorly endowed languages through digital tools and Artificial Intelligence. By doing so, we would like to encourage and help young Africans, who are willing to learn and/or teach their mother tongues, and therefore build a new generation of Africans aware of the importance and challenges of appropriating the languages and cultures of the continent. Another purpose of NTeALan's work is to provide local researchers and companies with data which could help them improve the quality of their services and work, hence building open-source African languages resources is one of our core projects.

2.2. NTeALan's approach: collaboration-based model

Our approach is exclusively based on the collaboration model (Holtzblatt and Beyer, 2017). We would like to allow African people to contribute to the development of their own mother tongues, under the supervision of specialists and academics of African languages. Our model involves setting up several communities: a community of speakers of these languages, a community of native specialists (guarantors of traditional, cultural and linguistic knowledge), a community of academics specialized in African linguistic technologies and a community of social, institutional and public partners. Grouped by languages, these communities work together with the same goal: building linguistic and cultural resources useful for research, technological and educational needs.

This approach applies to all NTeALan's internal projects, especially to the language resources platforms, as well as their representation.

3. NTeALan's language resource platforms

Our language resource platforms are divided into three parts: one independent architecture and two dependent architectures. The independent architecture serves not only the two others but also all NTeALan's projects as illustrated in figure 1.

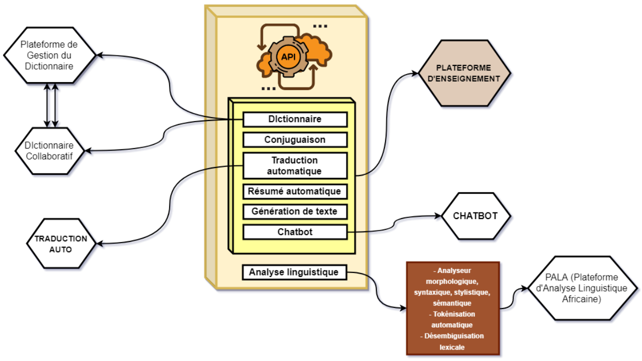
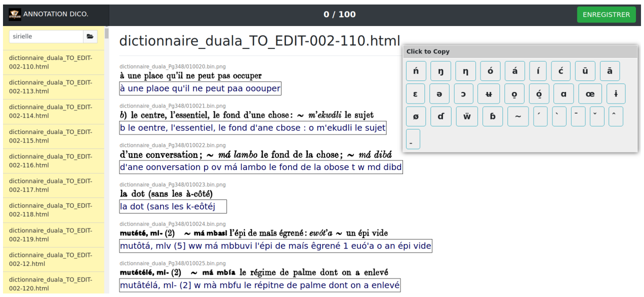

Plateforme de Gestion du Dictionnaire
DIctionnaire Collaboratif
TRADUCTION AUTO
API
...
...
DIctionnaire
Conjuguaison
Traduction automatique
Résumé automatique
Génération de texte
Chatbot
Analyse linguistique
PLATEFORME D'ENSEIGNEMENT
CHATBOT
- Analyseur morphologique, syntaxique, stylistique, sémantique
- Tokénisation automatique
- Désembiguisation lexicale
PALA (Plateforme d'Analyse Linguistique Africaine)

Figure 1: NTeALan APIs and service infrastructures

The three architectures are the fruit of two upstream processes depending on the input type (PDF files or images). The first process involves digitization and the second serialization:

- **digitization**: dictionaries in paper or digital format like PDF, TIFF, PNG by OCR (Optical Character Recognition) are digitized with Deep learning (Breuel, 2008); we annotate them to improve the OCR (see figure 2); each article constituents (featured word, translation, contextualization, conjugation, dialect variant, etc.) are automatically detected, extracted and xml-ized in XND (XML NTeALan Dictionary) format afterwards.
- **serialization**: dictionaries in an external format (toolbox, XML, TEI, LMF) are automatically serialized in XND format, using our internal NLP tools[2].

In both cases, we start with a paper or digital dictionary and end up with a XML dictionary in XND format. The latter is the unique data entry format for our three architectures. It should be noted that the two processes described above are controlled by NTeALan linguists only. In future work they will be opened to non-member contributors.



dictionnaire_duala_TO_EDIT-002-110.html
dictionnaire_duala_TO_EDIT-002-113.html
dictionnaire_duala_TO_EDIT-002-114.html
dictionnaire_duala_TO_EDIT-002-115.html
dictionnaire_duala_TO_EDIT-002-116.html
dictionnaire_duala_TO_EDIT-002-117.html
dictionnaire_duala_TO_EDIT-002-118.html
dictionnaire_duala_TO_EDIT-002-119.html
dictionnaire_duala_TO_EDIT-002-12.html
dictionnaire_duala_TO_EDIT-002-120.html

dictionnaire_duala_TO_EDIT-002-110.html

dictionnaire_duala_Pg348/010020.bin.png
à une place qu'il ne peut pas occuper
à une plaoe qu'il ne peut paa ooouper

dictionnaire_duala_Pg348/010021.bin.png
b) le centre, l'essentiel, le fond d'une chose: ~ *m'ekwáli* le sujet
b le oentre, l'essentiel, le fond d'ane cbose : o m'ekudli le sujet

dictionnaire_duala_Pg348/010022.bin.png
d'une conversation; ~ *má lambo* le fond de la chose; ~ *má dibá*
d'ane oonversation p ov má lambo le fond de la obose t w md dibd

dictionnaire_duala_Pg348/010023.bin.png
la dot (sans les à-côté)
la dot (sans les k-eôtéj

dictionnaire_duala_Pg348/010024.bin.png
mutété, mi- (2) ~ **má mbasi** l'épi de maïs égrené: *ewót'a* ~ un épi vide
mutôtá, mlv (5] ww má mbbuvi l'épi de maís êgrené 1 euó'a o an épi vide

dictionnaire_duala_Pg348/010025.bin.png
mutétélé, mi- (2) ~ **má mbía** le régime de palme dont on a enlevé
mutâtélá, ml- (2] w mà mbfu le répitne de palme dont on a enlevé

Figure 2: NTeALan dictionaries annotation platform based on Ocropy tool and used to train Deep learning model for OCR. This platform is under license on Creative Commons BY-NC-SA 3.0 license: (`http://dico-edit.ntealan.net`)

Figure 2 shows an example of annotation (from the bilingual Duala-French dictionary) performed by NTeALan's members.

3.1. Independent architecture

The independent platform is a web-based REST API platform. It can also be called lexicographical resources management database. Built to be simple and accessible, this web application stores and distributes all the lexicographic resources resulting from the collaborative work done by NTeALan's communities members and external contributors.

The independent architecture uses our internal NLP tools to manage the XND file format in order to give users easy

Department of Linguistics and African Literature at the University of Douala.

[2] These include tokenizers, lemmatizers, text parsers and lexical disambiguation tools used for processing noisy lexicographic corpora.

access to their contributions (see section 4.). The operations listed in table 3.1. are authorized in open access for each type of user.

Operations	NTeALan's users	Native speakers community	Scientific experts
manage dictionary	yes	no	yes
manage article	yes	yes	yes
validations	no	yes	yes
cultural media	yes	yes	no
comments	yes	yes	yes

Table 1: Users' privileges for each operation in NTeALan's REST API

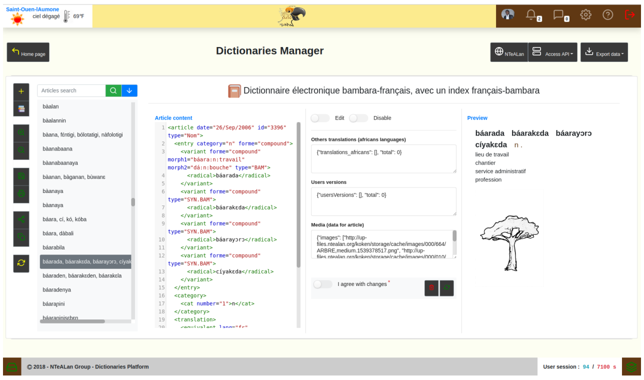
This architecture is hosted at `https://apis.ntealan.net/ntealan/dictionaries` and is accessible under the Creative Commons BY-NC-SA 3.0 license. The access rights, for each type of user, is described in table 3.1..

3.2. Dependent architectures

Dependent architectures are web platforms which use the data stored in common REST API database (Independent platform), the latter are enriched by contributors. They can also perform the operations described in table 3.1. through their web interface.

3.2.1. Dictionaries management platform

As a web platform, the dictionaries management platform is a graphical management version of the REST API platform. It allows NTeALan members (users) to manage dictionaries, articles, users, users comments, access requests and cultural resources.



```
<article date="26/Sep/2006" id="3396"
type="Nom">
  <entry category="n" forme="compound">
    <variant forme="compound"
morph1="báara:n:travail"
morph2="dá:n:bouche" type="BAM">
      <radical>báarada</radical>
    </variant>
    <variant forme="compound"
type="SYN.BAM">
      <radical>báarakɛda</radical>
    </variant>
    <variant forme="compound"
type="SYN.BAM">
      <radical>báarayɔrɔ</radical>
    </variant>
    <variant forme="compound"
type="SYN.BAM">
      <radical>cíyakɛda</radical>
    </variant>
  </entry>
  <category>
    <cat number="1">n</cat>
  </category>
  <translation>
    <equivalent lang="fr"
```

Others translations (africans languages)

{"translations_africans": [], "total": 0}

Users versions

{"usersVersions": [], "total": 0}

Media (data for article)

{"images": ["http://up-files.ntealan.org/koken/storage/cache/images/000/664/ARBRE,medium.1539378517.png", "http://up-files.ntealan.org/koken/storage/cache/images/000/010/

I agree with changes *

báarada báarakɛda báarayɔrɔ cíyakɛda n .
lieu de travail
chantier
service administratif
profession

© 2018 - NTeALan Group - Dictionaries Platform

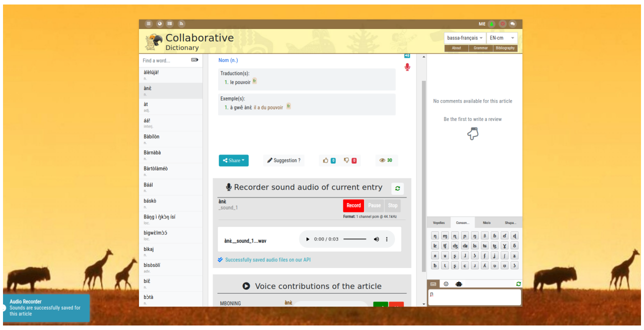
User session : 94 / 7100 s

Figure 3: Dictionaries management platform for managing multi-modal and multilingual lexicographical resources in African languages. This platform is under NTeALan's license: (`https://ntealan.net/dictionaries-platform`)

Unlike the two above-mentioned platforms, this is not an open-source platform. It can be used strictly by NTeALan's communities, in a direct collaboration between the linguistics team members and other association members.

3.2.2. Collaborative dictionary platform

The collaborative dictionary[3] is also a web platform (see figure 4) which enriches the lexicographic resources from the REST API. It gives NTeALan's communities members (see section 2.2.), more precisely native speakers and African languages experts, the opportunity to build, in a collaborative approach, resources like lexicons[4], illustration of cultural phenomenon, sounds and videos (recording process) based on semantic information provided by article written in their native languages. These shared resources are stored and freely available for all contributors through our REST API.



Nom (n.)

Traduction(s):
1. le pouvoir fr

Exemple(s):
1. à gwě ànὲ il a du pouvoir fr

No comments available for this article

Be the first to write a review

Recorder sound audio of current entry

ànὲ
_sound_1

Format: 1 channel pcm @ 44.1kHz

ànὲ__sound_1...wav 0:00 / 0:03

Successfully saved audio files on our API

Voice contributions of the article

Audio Recorder
Sounds are successfully saved for this article

Figure 4: Collaborative dictionaries for sharing multi-modal and multilingual lexicographical resources in African languages. This platform is under Creative Commons BY-NC-SA 3.0 license: (`https://ntealan.net`)

4. NTeALan language resources and representation

Most of our dictionaries resources are old bilingual dictionaries (from linguists' work) found on the web as open-source or under Creative Commons BY-NC-SA 3.0 license. The references to the original sources and to the NTeALan's versions are provided on all our platforms from where they can also be consulted.

4.1. African language resource dictionaries

We currently host and share 7 bilingual dictionaries[5] on our REST API. Although the number of entries to date is

[3] This project was born following the research work of Elvis Mboning at the University of Douala and University of Lille 3 (Master thesis): (Mboning, 2016) and (Mboning, 2017). We can cite other related work to this field like (Assoumou, 2010), (Mangeot and Enguehard, 2011), (Vydrin et al., 2016), (Maslinsky, 2014), (Nouvel et al., 2016), etc.

[4] To this aim, we built another platform to manage lexicographic resource: [https://ntealan.net/dictionaries-platform].

[5] Although the first versions are bilingual, these dictionaries are meant to be multilingual, with priority being given to translation in all the foreign languages spoken in Africa.

still relatively limited (from 3 to 11,500 entries), a growing community is participating daily in their filling. Table 4.1. shows the current statistics on the resources managed by our API.

Language resources	Entries	Entries contrib.	Media contrib.
Bambara-French	11487	1	1
Yemba-French	3031	2	90
Bassa-French	427	5	5
Duala-French	191	5	0
Ghomala-French	16	1	0
Ngiemboon-French	3	2	1
Fulfulde-French	0	0	0

Table 2: State of the art of NTeALan language resources currently saved in the REST API

Even if the current resources are insufficient and cover only 7 sub-saharan languages, we are nevertheless satisfied with the craze that is beginning to appear within the communities of users behind our platforms. However we would like to determine whether our different infrastructures fit with the resources produced, the load of connected users and the users needs. Once we have completed the tests on the platform, the next steps will be generalizing the model to the other African languages included in our dictionaries.

4.2. Description of NTeALan's XML format

Each lexical resource management platform has its own model for structuring and presenting data (sample of (Mangeot, 2006) and (Benoit and Turcan, 2006)). The XML format (mainly TEI and LMF XML standards) is today a reference choice for structuring linguistic, lexicographic and terminographic data. However, it turns out that these standards are not often adapted to represent and describe African languages. Indeed, several linguistic phenomena such as the concept of nominal class, the management, translation and localisation of dialect variants, and the notion of clicks are not explicitly treated, despite all the needs expressed with regard to the matter[6].

After analyzing the structure of a Bantu language from Cameroon (Yemba, spoken in West region), we decided to define a proprietary XML structuring model, whose structure was inspired by the 4 major families of African languages, namely the Afro-Asian family, the Niger-Kordofan family, Nilo-Saharan family and the Koisan family. Three principles guided our choice: representation, simplification and extensibility:

- **representation**: this principle aims at describing language data at the smallest morpho-syntactic level i.e. word components (prefix+radical+suffix) and phrase components like class accord (1/2, 2/4, 5/7).

- **simplification**: we try to choose XML tag names and international languages that are easily comprehensible for the research communities. Also, we decided to use a linear XML representation, with less parents and more children in the same parent node.

- **extensibility**: we would like to give external contributors the possibility to extend our main XML structure by adding new nodes (children or parent nodes), depending on the element to represent.

We design our *core-node* lexicographic data with a root node called `<ntealan_dictionary>`, which is divided into two subnodes: `<ntealan_paratexte>` and `<ntealan_articles>`. `<ntealan_paratexte>` describes the metadata about the version(s) of the document (context of the dictionaries production, source description of the original authors and target description of the XML VERSION). `<ntealan_articles>` describes all the dictionary articles (`<article>`).

Each article has its own subnodes: `<entry>` (dialect variant currently processed), `<category>` (grammatical categor(y/ies) associated to the dialect variant(s)), `<translations>` (translations associated to the dialect, `<examples>` (contextualisation of the dialect variants). Figure 5 illustrates this data representation.

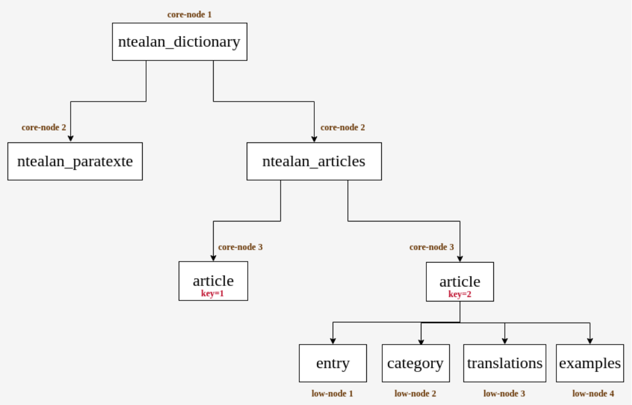
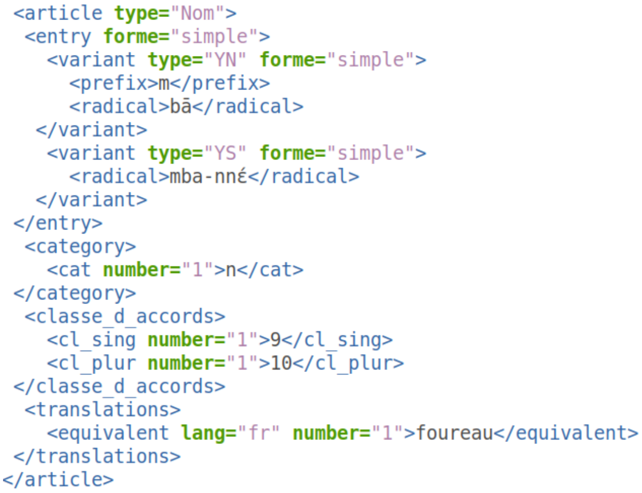


Figure 5: NTeALan dictionaries XML representation

The extension of the article structure by contributors is only possible in *low-node*, as shown in figures 5, 6 and 7, which means that the article model can be updated at each node level (referred to by an id).

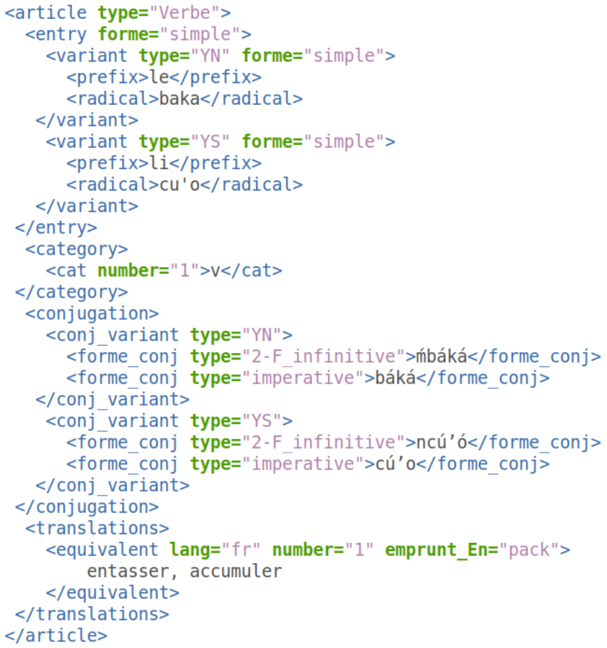
Our XND format is not intended to be standardized to serve as a reference. On the contrary, it is used as intermediate format, required by our internal NLP tools and by well-known standardized formats. Indeed once the external formats are serialized in XND, we have the possibility to convert the data into other formats such as those of the TEI and LMF dictionaries. These features will be available at the API level soon.

[6] Note that it is nonetheless possible in these standards to add new formalism (tags and attributes) in addition to existing classes.



```
 <article type="Nom">
  <entry forme="simple">
    <variant type="YN" forme="simple">
      <prefix>m</prefix>
      <radical>bā</radical>
   </variant>
    <variant type="YS" forme="simple">
      <radical>mba-nnɛ́</radical>
   </variant>
 </entry>
  <category>
    <cat number="1">n</cat>
 </category>
  <classe_d_accords>
    <cl_sing number="1">9</cl_sing>
    <cl_plur number="1">10</cl_plur>
 </classe_d_accords>
  <translations>
    <equivalent lang="fr" number="1">foureau</equivalent>
 </translations>
</article>
```

Figure 6: Sample Xmlisation of nouns article *mbā* extracted from the Yemba-French dictionary



```
<article type="Verbe">
  <entry forme="simple">
    <variant type="YN" forme="simple">
      <prefix>le</prefix>
      <radical>baka</radical>
   </variant>
    <variant type="YS" forme="simple">
      <prefix>li</prefix>
      <radical>cu'o</radical>
   </variant>
 </entry>
  <category>
    <cat number="1">v</cat>
 </category>
  <conjugation>
    <conj_variant type="YN">
      <forme_conj type="2-F_infinitive">ḿbáká</forme_conj>
      <forme_conj type="imperative">báká</forme_conj>
   </conj_variant>
    <conj_variant type="YS">
      <forme_conj type="2-F_infinitive">ncú’ó</forme_conj>
      <forme_conj type="imperative">cú’o</forme_conj>
   </conj_variant>
 </conjugation>
  <translations>
    <equivalent lang="fr" number="1" emprunt_En="pack">
        entasser, accumuler
    </equivalent>
 </translations>
</article>
```

Figure 7: Sample Xmlisation of verbs article *lebaka* extracted from Yemba-French dictionary.

5. Problems encountered and future challenges

The implementation of these first platforms enabled us to take note of the main challenges. In upcoming years, we will focus on these issues, enriching our platforms and trying to improve them for future deadlines.

5.1. Problems encountered

We are currently facing two main problems with the NTeALan platforms:

- the first is the low number of contributors and the insufficient IT resources. The staff do not have all the specialists needed (in NLP, NLU and African languages) for the targeted goals and great ambitions. The current work is mainly carried out by 4 active members of the association. Regarding IT resources, we do not have enough robust IT infrastructures (servers, field tools, etc.) as required by such a research work on African languages.

- the second is the lack of funding to carry out our research activities with respect to the development of NLP and NLU tools. Our funding mainly comes from the contributions of the association members, which is not enough in the light of our current ambitions.

5.2. Further challenges

Our ambitions are great and will require more staff (language specialists) and financial resources. We would like to:

- Above all, encourage the greatest number of specialists in African languages and cultures from various African countries and in the whole world, to join our association because together we can easily take up challenges.

- Find funding from private and public institutions, businessmen, companies, who can support our research work and the continuous development of our applications for the teaching of poorly endowed African languages.

- Enrich and improve all existing platforms and open them up more to the scientific community and to speakers of the languages included. We will first of all focus on : the autonomous platform for language and culture teaching, the conversational Agent Assistant for Language Teaching and the Virtual cultural museum for safeguarding the African socio-cultural inheritance.

- Strengthen our partnerships with social and cultural African institutions, universities, research laboratories and companies specialized in our research areas. The aim is to create communities of experts in linguistics, technological and cultural issues throughout the continent.

De Schryver (De Schryver, 2010, p.587) already wondered about the specifics of electronic lexicography in the future in these terms: "The future of lexicography is digital, so much is certain. Yet what that digital future will look like, is far less certain.". This work clearly shows that the collaboration-based model, coupled with robust NLP platforms, could give meaning to the future nature of electronic lexicography in Africa.

6. Conclusion

In this article, we described NTeALan platforms and its XND data representation, and we showed how essential an association is nowadays, for the construction of good linguistic and lexicographic resources and tools for endowed African languages. We lead, internally with our academic partners (the language and African literature department of the University of Douala and the ERTIM team of INALCO (France)), numerous research activities in Artificial Intelligence, NLP, and NLU, in order to contribute to the industrialization of African languages. It is obvious that a lot remains to be done, however the first results of our study have proven to be very useful for our applications (the conversational agent NTeABot, the learning platform, the translation platform, etc.) and can be used by other researchers: this includes data (in different common formats like XML, TEI, LMF, XND) and tools. We are convinced, as Tunde Opeibi (Tunde, 2012, p.289) already said, that "the linguistic diversity in Africa can still become the catalyst that will promote cultural, socio-economic, political, and technological development, as well as sustainable growth and good governance in Africa."

7. Acknowledgements

These platforms was developed by Elvis Mboning (REST API and Dictionaries Management platform), Daniel Baleba (Collaborative dictionaries) and Jean-Marc Bassahak (Web Design and interfaces). NTeALan's projects are actually supported by the Ministry of Post and Telecommunication of Cameroon, the Department of linguistics and African literature of the University of Douala (Cameroon), the INALCO's ERTIM research team and Fractals system (France). We can also cite: Professor Jules Assoumou, Thanks to Christian Bonog Bilap, Marcel Tomi Banou, Ntomb David, Ntomb Nicolas, Théophile Kengne, Yves Bertrand Dissake, Juanita Fopa, Damien Nouvel and to the other contributors.

8. Bibliographical References

Abate, S. T., Melese, M., Tachbelie, M. Y., Meshesha, M., Atinafu, S., Mulugeta, W., Assabie, Y., Abera, H., Ephrem, B., Abebe, T., Tsegaye, W., Lemma, A., Andargie, T., and Shifaw, S. (2018). Parallel Corpora for bi-lingual English-Ethiopian Languages Statistical Machine Translation. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3102–3111, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Assoumou, J. (2010). *Enseignement oral des langues et cultures africaines à l'école primaire*. Éditions Clé, Yaoundé, Cameroun, 1st edition.

Benoit, J.-L. and Turcan, I. (2006). La TEI au service de la transmission documentaire ou de la valorisation des richesses patrimoniales : le cas difficile des dictionnaires anciens.

Breuel, T. M. (2008). The OCRopus open source OCR system. *Proc.SPIE*, 6815.

De Schryver, G.-M. (2010). State-of-the-Art Software to Support Intelligent Lexicography. *ResearchGate*, page 16.

Don, O. (2010). African languages in digital space. *HSRC Press*, page 168.

Holtzblatt, K. and Beyer, H. (2017). 7 - Building Experience Models. pages 147–206, January.

Mangeot, M. and Enguehard, C. (2011). Informatisation de dictionnaires langues africaines-français. In *journées LTT 2011*, page 11.

Mangeot, M. (2006). Dictionary building with the jibiki platform. In Cristina Onesti Elisa Corino, Carla Marello, editor, *Proceedings of the 12th EURALEX International Congress*, pages 185–188, Torino, Italy, sep. Edizioni dell'Orso.

Maslinsky, K. (2014). *Daba: a model and tools for Manding corpora*.

Mboning, E. (2016). De l'analyse du dictionnaire yémba-français à la conception de sa DTD et de sa réédition sur support numérique. Mémoire Master 1, Université de Lille 3.

Mboning, E. (2017). Vers une métalexicographie outillée : conception d'un outil pour le métalexicographe et application aux dictionnaires Larousse de 1856 à 1966. Mémoire Master 2, Université de Lille 3.

Nouvel, D., Donandt, K., Auffret, D., Maslinsky, K., Chiarcos, C., and Vydrin, V. (2016). Resources and Experiments for a Bambara POS Tagger. *Intra Speech*, page 14.

Ruthven, R. (2005). The African Association for Lexicography: After Ten Years. *Lexikos journal*, page 9.

Tadadjeu, M. (2004). African Language Needs in Information and Communication Technology (ICT). page 9.

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

Vydrin, V., Rovenchak, A., and Maslinsky, K. (2016). Maninka Reference Corpus: A Presentation. In *TALAf 2016 : Traitement automatique des langues africaines (écrit et parole). Atelier JEP-TALN-RECITAL 2016 - Paris le*, Paris, France, July.

9. Language Resource References

De Pauw, Guy and Waiganjo Wagacha, Peter and de Schryver, Gilles-Maurice. (2009). *The SAWA corpus: a parallel corpus English - Swahili*.

Mboning, Elvis and NTeALan contributors. (2017). *NTeALan lexicographic African language resources: an open-source REST API*. NTeALan Project, distributed via NTeALan, Bantu resources, 1.0.

Vydrin, Valentin and Rovenchak, Andrij and Maslinsky, Kirill. (2016). *Maninka Reference Corpus: A Presentation*. Speecon Project, distributed via ELRA, Madingue resources, 1.0, ISLRN 613-489-674-355-0.