

## *Constitution et Usage de Corpus pour Langues Terminologiquement Peu Documentées*

---

Abibatou Diagne

[abibatou.diagne@ucad.edu.sn](mailto:abibatou.diagne@ucad.edu.sn)

*Université Cheikh Anta Diop de Dakar, Sénégal*

**Résumé** - Dans les travaux de Diagne (2018), Diagne et Kébé (2018), la nécessité de travailler à la constitution de corpus à des fins terminologiques a été soulignée. Quels sont les procédés à disposition, lorsque la langue étudiée est lexicalement limitée dans les domaines techniques et scientifiques ? C'est à cette question que le présent article propose de répondre.

Un corpus à visée terminologique peut être créé en utilisant deux méthodes. La première est axée sur la production d'unités terminologiques. Elle consiste à établir des listes d'unités lexicales spécialisées. La création d'un corpus peut également - c'est la seconde méthode - se faire au niveau discursif à travers la recherche et la compilation de productions en lien avec la thématique choisie. Des exemples seront alors pris des travaux effectués en terminologie dans plusieurs domaines, dans notre cas, juridique et médical en langue wolof.

La notion de langue terminologiquement peu documentée renvoie à toute langue dont l'usage dans les sciences, la technologie ou d'autres domaines de spécialité n'est pas très développé. Il y a deux caractéristiques majeures de ces langues en contexte africain. La première est que le cadre d'émergence des unités lexicales spécialisées est assez éloigné du contexte sociolinguistique de la langue à documenter. La seconde est qu'une langue terminologiquement peu documentée a un usage essentiellement oral. Même si la langue dispose de sa propre transcription ou est codifiée, il n'existe pas de littérature scientifique spécialisée comparable à celle d'une langue à forte tradition écrite. Les limites de ces langues sont néanmoins uniquement d'ordre quantitatif, elles ne remettent pas en cause la capacité qu'elles ont à exprimer des domaines de spécialité.

**Mots-clés** : corpus, terminologie, langues en Afrique, aménagement linguistique, traduction.

**Abstract** - In our previous work Diagne (2018) Diagne et Kébé (2018), we insisted on the need to work in the sense of gathering corpora for terminological purposes. What procedures are available when the language in study is lexically limited in technical and scientific fields? This is the question that this study proposes to answer.

A corpus can be created using two methods. The first focuses on the production of terminological units. From this point of view, it consists in establishing lists of specialized lexical units. It is possible to carry out a translation exercise from a foreign language into the language to be terminologically documented. The creation of a corpus can also be done at the discursive level through the research and compilation of discursive productions in connection with the chosen theme. On this specific point, sample of compilations will be taken from work carried out in legal and medical terminology in the Wolof language.

The notion of a terminologically poorly documented language refers to any language whose use in science, technology or other fields of specialization is not very developed. There are two major characteristics of these languages in the African context. The first one is that the framework in which emerge specialized lexical units is quite distant from the sociolinguistic context of the language to

document. In addition, a terminologically poorly documented language is used primarily oral. Even if the language has its own transcription or is codified, there is no scientific, specialized literature comparable to that of an official language with a strong written tradition. The limits of these languages are quantitative, they do not call into question the ability of these languages to express areas of specialization.

**Key-words:** corpus, terminology, languages in Africa, language planning, translation.

**Tënk (Wolof)** - Ci yenn ci suñuy gëstu yi weesu, Jaañ (2018), Jaañ ak Kebe (2018), biraloon naa ni warees nañu liggéey ci dajale ay bëyit ak ay njangale (koopis) ngir mën a liggéey ci baati-xam-xam yi ci làkki réew mi. Naka la ñuy def ba ñuy xalaat ci xarala ak yenn xam-xam yu kawé yi su fekke ni làkki réew mi amu ñu baati-xam-xam yi jëm ci wàll yooyu ? Laaj bii la gëstu bi sàmp te war ci tontu.

Mënees na dajale koopis jaarale ko ci ñaari yoon. Bi ci jëkk mooy tegg ay baati-xam-xam maanaam boole leen ba nu tollu ci ab lim. Loolu mën nañu ko def suñu jëlee aw làkku

bitim-réew daldi ko tekki ci suñuy làkki réew. Beneen yoonu defar koopis mooy boole ay mbind ak i kàddu yu tukke ci ay boroomi xam-xam yuy yëngu ci wenn wàll. Misaal yu bare yu ñuy jël jème ko ci koopis yi, ñoo ngi leen di jële ci wàllu pacc ak yoon.

Li aju ci làkk yi bariwul baati-xam-xam, du nuy faral di leen di jëfandikoo ci wàllu xeltu ak xam-xamu xarala. Am nañu ñaari màndarga. Bi ci jëkk mooy ne baati-xam-xam yi tukkee bitim-réew fa ñuy jòge ak nañu leen di tégge dafa soreyantoo lool ak ni ñuy tégge ak a jëfandikoo suñuy lakki réew, nga xam ne du ay làkk yu ñuy faral di bind. Loolu sax mooy ñaarelu màndarga bi ; lakki réew mi am nañu seeni mbind ak lepp waaye, du nu leen di jëfandiko lu bare ci wàllum xeltu ak xam-xam yu kawé yi. Ci gàttal daal, làkki réew mi, mënees na cee amal ay xeltu, def ci ay gëstu yu xóot waye fii ñu tollu nii, mbind mi ak gëstu yi deme wu ñu noonu.

## INTRODUCTION

Les études qui se fondent ou qui portent sur le corpus sont très variées. Le corpus, en linguistique, est un outil d'appoint de recherche, de repérage et d'attestation de faits langagiers, s'il n'est pas considéré comme une approche théorique à part entière selon un modèle hypothético-déductif.

In a corpus-driven approach the commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence. (Tognini-Bonelli, 2001 : 84)

C'est ainsi que dans une approche corpus driven, le corpus sert de base empirique à partir de laquelle l'on extrait des données qui ne sont sous-tendues par aucune affirmation apriorique.

L'usage du corpus dans d'autres domaines tels que les arts ou encore la littérature s'inscrit dans une démarche de recherche, d'analyse de régularité en vue d'interpréter des œuvres. Le corpus est un ensemble dont les éléments constitutifs partagent des caractéristiques. Il jouit d'un ancrage certain dans les sciences humaines et sociales.



En linguistique, le corpus est inséparable de la notion de données langagières ou encore de ce qui s'érige en discipline autonome : la linguistique de corpus. Pour une recherche scientifique donnée portant sur des aspects phonologiques, morphologiques, sémantiques, syntaxiques, sociolinguistiques, etc., le corpus tend à l'exhaustivité (notion critiquée qui sera abordée), mais l'on est plutôt en face de données constituées et arrêtées en un temps et en un lieu afin d'établir des conclusions et de diffuser des résultats. L'autre aspect majeur du corpus en tant que données langagières est la notion de représentativité (également abordée). Un corpus est représentatif de quoi ? Il est représentatif en synchronie et en diachronie des aspects linguistiques et langagiers abordés : le français au Sénégal ; les variations de l'anglais en Gambie ; l'évolution morphophonologique d'une langue d'un siècle à un autre, etc. De ce point de vue, il est davantage question d'échantillonnage Habert et *al* (1997) c'est-à-dire un échantillon prélevé de l'existant que d'exhaustivité (toutes les données possibles constituées pour les besoins d'une recherche spécifique).

L'idée d'échantillonnage a forgé la linguistique de corpus. À la question de savoir ce qu'est un corpus le linguiste britannique John Sinclair le définit comme : « une collection de données langagières sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage » Sinclair (1991). C'est également dans la tradition britannique que l'on retrouve les approches de 'corpus driven' (analyse de données sans théories aprioriques), et de 'corpus based' (le corpus est un outil d'appoint qui atteste ou valide). Il faut toutefois rappeler que le développement de la linguistique de corpus est fortement impulsé par l'ordinateur en tant qu'outil informatique à grande capacité mémorielle et de stockage, comme le note Renouf (1987).

Using computer techniques extensively to build a substantial corpus of English text, to analyse it objectively and exhaustively, and to capture the insights gained from that analysis in an electronic data base has moved the science of lexicography into a new phase. The power of the machine to store, search, classify, sort and otherwise manipulate language data in vast quantities liberates the study of language (...). (Renouf, 1987 : 61).

L'utilisation des corpus en terminologie sert à la recherche de termes nouveaux ou attestés. Ce travail est davantage facilité par les moyens informatiques et la semi-automatisation du repérage. Les corpus servent de cadre ou de contexte à partir duquel sont extraites des informations. La constitution de listes de terminologies peut elle-même se présenter comme un corpus.

## 1. Corpus et compilation de terminologies

La compilation se fait selon plusieurs objectifs. En linguistique, l'usage continu, dans le cadre de recherches, d'attestations de validations empiriques ou encore de références, a contribué à l'émergence de la linguistique de corpus. Elle occupe en linguistique, selon les traditions, une position quelque peu floue. En effet dans le cadre empiriste britannique, c'est une discipline autonome qui peut faire l'objet d'une théorisation. Alors que pour d'autres linguistes de tradition francophone, elle reste un outil d'appoint et de vérification d'hypothèses.

### 1.1. Quelques principes de compilation

Les principes de compilation de corpus sont fonction de l'objectif visé. Il y a toutefois un certain nombre d'aspects communs à tout corpus regroupé à des fins scientifiques. L'un des points essentiels est l'information ou la documentation du corpus également appelée métadonnées. Quelle que soit l'origine ou le but de la recherche sur corpus, il est nécessaire de donner le plus d'informations possibles relatives à son auteur, sa datation, les sources, les conditions de constitution telles que le processus, les enregistrements la transcription ou retranscription. En matière de corpus, l'anonymisation est un point crucial indissociable de questions éthiques et juridiques.

L'impression d'accessibilité de données foisonnantes produites sur le net est souvent trompeuse, dès lors que de nombreuses publications sont protégées par des droits d'auteur. En milieu scientifique, les auteurs sont pour la plupart ouverts à partager des données textuelles ou sonores qu'ils ont déjà exploitées tant que cela reste dans le cadre scientifique et qu'il n'y a pas d'objectif de commercialisation. Dans tous les cas de figure, il apparaît judicieux de demander l'autorisation d'usage partiel ou total de travaux aux auteurs concernés ou à leurs ayants droit.

Pour ce qui est des types de corpus (oral et écrit), certaines spécifications dues à leur caractère sont nécessaires. Dans le cas des langues terminologiquement peu documentées, la rareté (inexistantes ou peu accessibles) de données écrites et parfois l'ancienneté de celles qui existent font qu'il y a un besoin de les adapter aux impératifs de traitement automatique et de formatage de textes. Ce sont là quelques aspects techniques, mais nécessaires pour un usage optimal. La fragilité de ces documents impose souvent un nouveau travail de dactylographie. Si la qualité du document le permet, il est possible de faire usage de l'OCR (Optic Character recognition) pour passer d'un document textuel numérisé ou photographié à une suite de caractères. Afin de donner davantage de visibilité aux données textuelles des langues terminologiquement peu documentées, en particulier sur le Net, la réécriture ou encodage du corpus au format XML est



d'une grande utilité. C'est un point qui a le mérite d'élargir l'accessibilité du corpus à une plus grande communauté scientifique. Le caractère spartiate de l'écriture informatique fait que les chercheurs en sciences humaines sont peu enclins à les utiliser alors qu'il existe des logiciels peu compliqués à manipuler. Il faudrait toutefois rappeler que l'encodage XML demeure un choix en vue d'un élargissement d'accès (sauf pour les pages HTML mises au format XML), mais non pas une exigence pour le traitement et l'analyse du corpus lui-même. Ce n'est pas le cas du formatage du texte (passage d'un PDF ou d'un document DOCX à un format texte brut). Tous les procédés mentionnés ne doivent pas occulter le passage en revue des données textuelles regroupées afin de procéder à un nettoyage manuel des coquilles, scories et autres erreurs orthographiques, notamment lorsqu'il s'agit de documents passés à la reconnaissance optique de caractères (OCR).

Le corpus oral qui nous intéresse tout particulièrement ne sera jamais que de l'oral retranscrit pour un meilleur traitement, mais surtout pour rester le plus fidèle possible aux données. Le seul travers dans lequel, il est possible de tomber avec l'oral retranscrit, est le risque de surinterprétation puisque le découpage syntaxique se révèle souvent subjectif. La longueur des pauses est interprétée, selon l'intuition prosodique de la personne qui retranscrit, comme une virgule ou un point, mais la notion même de phrase à l'oral demeure discutable. Comme pour l'écrit, des logiciels de transcription de l'oral sont disponibles. L'alignement des segments de retranscription avec le document sonore offre une meilleure analyse en contexte. Aussi bien pour l'écrit que pour l'oral l'analyse de corpus à l'aide d'un concordancier permet un travail plus rapide. La majorité des concordanciers sont des gratuits qui ne traitent, pour la plupart, que le format texte brut. Le volume des corpus oraux après retranscription est à tempérer, compte tenu des différentes annotations qui, enrichissent le corpus et donnent des informations utiles.

## 1.2. Usage des corpus en terminologie

La terminologie s'est développée autour d'une approche *corpus-based* (Sager, 1990) au cours des dernières décennies. Du fait de la spécialisation des données langagières, le volume des corpus utilisés en terminologie est nettement moins important que ce qui est utilisé en lexicographie qui dispose de bien plus de ressources comme le souligne de Schatzen (1996), à la suite de Sinclair (1991). Cet aspect pose les questions de représentativité et d'échantillonnage mentionné supra et qu'il nous faut donc revoir dans le cas spécifique de la terminologie. Un travail sur corpus médical d'une dizaine de millions d'occurrences est tout à fait envisageable lorsqu'un intérêt est porté à des langues telles que le français ou l'anglais. Pour une langue terminologiquement peu documentée, l'échantillon

sera bien plus réduit. Pour autant, il sera représentatif des pratiques et données langagières dans ce domaine, pour cette langue. La représentativité ne saurait donc être réduite à des considérations purement quantitatives. Il apparaît judicieux de considérer la représentativité quantitative proportionnellement à la disponibilité effective des productions techniques ou scientifiques dans le domaine étudié pour une langue donnée.

La recherche de termes sur corpus se fait de manière semi-automatisée à l'aide de concordanciers ou tout autre outil jugé approprié. Un concordancier donne des statistiques lexicales et une vue globale de la structure du corpus. Par ailleurs pour le repérage de termes, il y a la possibilité de dégager des paradigmes lexicaux avec une mise en évidence du cotexte ou de catégories lexicales.

L'usage de corpus en terminologie sert de référence pour la recherche de termes appropriés en traduction. C'est un procédé qui est assimilé à un travail de recherche en contexte qu'effectue le traducteur. C'est un aspect davantage tourné vers l'usage de la terminologie en milieu professionnel et s'apparente également à la recherche documentaire.

Certains travaux linguistiques ont mis en avant le fait qu'au-delà du repérage de termes et de patrons syntaxiques qui se présentent comme des unités complexes, le corpus comporte des modèles de métalangage qui participent à la définition des termes. C'est ce que note Pearson (1998) qui relève par ailleurs que le cadre métalinguistique du corpus fournit des définitions pouvant être réutilisées. C'est là un élément clé pour une compilation dans une perspective multilingue. Nous rappelons plus haut que le cadre d'émergence des unités terminologiques est souvent éloigné du cadre sociolinguistique des langues terminologiquement peu documentées. Ces éléments de métalangage participent au rapprochement et à une meilleure compréhension des unités extraites, ce qui évite, ce qui serait un transfert mécanique d'unités terminologiques d'une langue donnée vers une langue à documenter. En effet, l'unité terminologique est la représentation matérielle, verbale du concept en terminologie. Cela ne devrait pas pour autant occulter son parcours cognitif.

L'unité terminologique est nécessairement motivée et renferme une forte valeur cognitive. Certains contenus cognitifs nous ont permis, avec nos recherches sur la terminologie médicale wolof, de trouver des désignations appropriées sur la base de métaphores qui se rapprochent le plus de la culture de cette langue. La connaissance technique ou scientifique d'une unité terminologique de la langue source se présente comme une condition nécessaire pour trouver des équivalents vers une langue cible.



## 2. Productions d'unités terminologiques

Des listes de terminologies peuvent en elles-mêmes servir de corpus. En contexte africain, deux orientations existent dans les pratiques terminologiques : une orientation-documentation et une orientation-développement (Antia, 2000). L'orientation documentation rassemble des ressources disponibles dans la langue concernée parce que la communauté utilisatrice a développé une forte expérience sur un domaine de spécialité (pêche, agriculture, certains métiers de l'artisanat). L'orientation développement, pour sa part, est une approche dans laquelle il y a un vide ou une rareté des ressources (informatiques, domaines technologiques, etc.).

Il convient toutefois de souligner qu'il n'y a pas de frontière étanche entre ces deux orientations proposées par Bassey. La pêche artisanale côtoie la pêche industrielle, l'agriculture est à la fois paysanne et intensive...c'est dire que l'évolution des domaines fait qu'il est surtout question de recension d'unités puis de mise à jour ou d'enrichissement du lexique existant. Deux domaines illustrent mieux ce fait : le juridique et le médical. Pour le domaine juridique et plus spécifiquement celui de l'organisation politico-administrative, on trouve des unités presque chez toutes les communautés constituées (chez les Wolofs, la communauté lébou a érigé l'une des organisations les plus anciennes qui perdure jusqu'à présent). La retranscription des composants de ces organisations est une base de travail réutilisable. La pluralité de ces ressources juridiques a incité Abolou (2011), à la suite de Goulet (1992) et Koubi (1997), à parler « d'entropie systémique » avec un entrecroisement et une superposition de normes juridiques. Il en va de même pour le domaine médical où la médecine traditionnelle cohabite avec la médecine moderne.

Que cela soit dans un but de recension ou d'enrichissement, l'on note bien qu'il est nécessaire de partir d'un corpus dont la nature et l'exploitation influent sur les terminologies regroupées.

D'abord au niveau linguistique, un travail de terminologisation qui ne se fonde que sur les corpus disponibles dans la langue à documenter ne pose pas de difficultés majeures puisqu'il est question de répertorier un lexique constitué. Dans certains cas, quelques réadaptations morphophonologiques (variations lexicales) peuvent s'imposer. En plus de ces possibles réajustements formels, l'évolution sémantique des unités fera l'objet d'une attention particulière. Le travail d'enrichissement terminologique comporte une dimension traductive à la fois inter et intra linguistique. Ce sont deux systèmes linguistiques qui sont en présence avec l'objectif d'enrichir l'un sur le plan lexical. Des questions lexicales se poseront : morphosémantique du terme, emprunt, procédés néologiques. La traduction au sein même de la langue est un travail de reformulation (adaptation

du terme au caractère verbalisant d'une langue comme le wolof) qui fait appel à l'usage de tournures paraphrastiques et à une considération de règles syntaxiques.

Lorsque le corpus est compilé à partir de la langue à documenter, les connaissances véhiculées montrent que la culture du locuteur coïncide avec son expérience dans le domaine d'étude. Mais cela n'est parfois pas le cas lorsqu'il s'agit d'un corpus compilé dans une langue étrangère. Le défi majeur est de trouver un équivalent cognitif, culturel à même de restituer sur le plan linguistique, cette réalité étrangère. C'est sous cette optique que la terminologie culturelle de Diki-Kidiri (2008) trouve tout son sens. Pour Diki-Kidiri, la pratique terminologique en contexte africain, pour qu'elle soit réussie, doit mettre en avant l'adéquation des désignations avec le cadre culturel cible. Pour ce faire, il met l'accent sur la nécessité d'identifier les éléments représentatifs de la culture cible afin que ses locuteurs puissent se réapproprier une réalité étrangère.

En s'inscrivant dans cette démarche philologique, culturelle, Diki-Kidiri fait également un choix délibéré de mettre de côté la normalisation qui est un pan tout aussi important. En effet, compte tenu du caractère ardu de la recherche, de la constitution de corpus puis de la création de terminologies, la normalisation se présente comme un aboutissement et une organisation pratique et structurée d'un processus. Cette organisation peut être un glossaire, une base de données, un dictionnaire de lexique spécialisé évolutif au gré des productions scientifiques, techniques orales et textuelles.

### **3. Productions orales et textuelles en contexte**

L'une des particularités des langues terminologiquement peu documentées est qu'elles font l'objet d'un usage essentiellement oral dans les pratiques langagières courantes. Toute constitution de corpus doit nécessairement tenir compte de cet aspect fondamental.

Pour la terminologie médicale wolof, le corpus oral compilé dans le cadre de nos travaux de recherches doctorales est composé d'enregistrements sonores qui proviennent des journaux parlés de stations de radios privées et publiques du Sénégal. La durée de ces enregistrements va de deux à sept minutes, pour vingt extraits sonores, soit un corpus de vingt-mille occurrences après transcription. La thématique abordée est celle de la santé (une rubrique santé consacrée à une édition). C'est une thématique large, mais qui apparaît comme l'une des rares opportunités d'avoir à disposition un corpus oral qui traite d'aspects médicaux. Dans ce travail, nous avons porté un intérêt linguistique aux termes extraits, en travaillant, d'une part, sur les patrons syntaxiques de type collocation (nom-





adjectif, verbe-nom principalement). La manière de restituer ce lexique spécialisé en wolof a été l'autre point majeur de ce travail.

Un corpus oral aurait également pu être compilé sur la base des dialogues médecins-patients, ce qui aurait été davantage au cœur de la thématique médicale. Toutefois compte tenu de la soumission de questionnaires aux professionnels de la santé sur l'usage des langues nationales du Sénégal en milieu médical, les dialogues médecins-patients ont été mis de côté. C'est une ressource qui mérite considération parce qu'elle colle, au plus près, à la trame discursive du domaine médical, contrairement à la thématique de la santé qui inclut beaucoup d'éléments sociaux, institutionnels, administratifs. L'autre utilité de ce type de corpus est qu'il permettrait d'élargir l'extraction et l'analyse des unités terminologiques à des aspects sociolinguistiques. Au-delà des considérations morphosémantiques propres aux termes, il est possible de les envisager au sein de la structure discursive, notamment la construction du sens, les usages qu'en font les énonciateurs, le fonctionnement des termes dans le discours, leur circulation et implantation, entre autres aspects.

Un autre mode de constitution entre expert d'un domaine et terminologue donne matière à constituer un corpus oral. Cela se ferait dans un contexte assez contrôlé avec des questions à orientations ciblées. C'est à la fois une occasion d'en savoir un peu plus sur le champ d'étude, mais également de déterminer les points linguistiques sur lesquels axer une recherche.

Pour ce qui est des productions écrites, dans le domaine juridique en langue wolof, il faut relever le travail de traduction de la Constitution du Sénégal (Diouf et Fall, 2010). C'est un corpus qui n'est pas auto-traduit, ce qui permet de garder une certaine distance entre traduction et appréciation terminologique. Même s'il s'agit de la Constitution du Sénégal, celle-ci étant pour l'essentiel calquée sur la Constitution française, l'on note des écarts culturels auxquels les traducteurs ont dû faire face. La notion de « République » (les traducteurs procèdent par emprunt adapté sur le plan morphophonologique), « démocratie » (emprunt adapté), syndicat (emprunt adapté) sont quelques-uns de ces écarts. Dans le champ médical, la plupart des écrits à disposition en langue wolof sont des livrets publiés par l'OSAD. Elle est aujourd'hui l'une des ONG qui promeut les langues nationales sénégalaises et s'est spécialisée dans les domaines de l'éducation non formelle et de l'édition d'ouvrages en langues nationales. Ces livrets ont une orientation pratique qui souligne les besoins et problèmes récurrents en matière de santé communautaire. Pour les corpus médicaux écrits en langue étrangère, la constitution sur la base de la matrice lexicale (affixes et formants grecs et latins) est l'une des ressources les plus considérables.

## CONCLUSION

Cette étude a rappelé quelques principes de compilation de corpus ainsi que les difficultés propres à la constitution pour les langues dont le lexique spécialisé n'est pas très développé pour les domaines scientifiques et techniques. Les pratiques terminologiques en contexte africain requièrent ce travail en amont pour deux raisons : recenser la documentation existante puis aller dans le sens d'un enrichissement lexical. Les exemples pris du wolof pour les domaines juridique et médical témoignent de certains faits sociolinguistiques perceptibles dans beaucoup de pays africains. Il s'agit d'une langue véhiculaire qui concurrence une langue étrangère dans les pratiques langagières courantes (orales principalement). Pour autant elle n'a pas statut de langue officielle et ne développe pas de culture écrite forte dans les domaines d'intérêt socio-économiques ou scientifiques. La terminologie, tout en participant au développement du lexique de spécialité, se nourrit des corpus soit pour mieux cerner le cadre d'émergence des unités et leur adéquation pour une communauté locutrice, soit pour renouveler le stock lexical des domaines de spécialité en constante évolution.

## BIBLIOGRAPHIE

- Abolou, C-r. 2011, « Le discours juridique en Afrique noire. Terminologie et traduction du droit », *Revue française de linguistique appliquée*, vol. xvi, no. 1, pp. 17-31.
- Antia, B.. 2000, *Terminology and language planning*, Amsterdam/ Philadelphia: John Benjamins.
- Diagne, A. 2018, *La terminologie wolof dans une perspective de traduction et de combinatoire lexicale restreinte*, Thèse de Doctorat : Université de Lyon 2.
- Diagne, A, et Kébé, A.B. 2018, « Vers une socioterminologie médicale wolof ». *Lidil*, no 58. <https://doi.org/10.4000/lidil.5029>.
- Diki-Kidiri, M., et Edema Atibakwa Baboya. 2008, *Le vocabulaire scientifique dans les langues africaines: pour une approche culturelle de la terminologie*. Paris: Karthala.
- Diouf, A, et Fall, A. 2010. " Ndeyu Àtte Repibligu Senegaal." , Dakar.
- Goulet, J. 1992. « Quelques variations sur le modèle thermodynamique et le droit artificiel », In Bourcier, D. & Mackey, P., *Lire le droit : langue, texte, cognition*, Paris, LCDJ, 29-40.
- Koubi, G. 1997. « Des-ordre/s juridiques/s. », In CURAPP (Université de Picardie) (éd.), *Désordres*, Paris, PUF, 201-214.



- Pearson, J. 1998, *Terms in Context*, Amsterdam/ Philadelphia: John Benjamins Publishing.
- Renouf, A. 1987, « Corpus Development ». In *Looking up, An account of the Cobuild Project lexical computing and the development of the Collins COBUILD English Language Dictionary*, édité par John Sinclair, 1-40. Londres: Collins ELT.
- Schaetzen, C. 1991, « Corpus et terminologie : constitution de corpus pour la confection des dictionnaires. », *Revue française de linguistique appliquée*.
- Sinclair, J. 1991, *Corpus, Concordance, Collocation*. Oxford University Press, Oxford University Press.
- Tognini-Bonelli, E. 2001, *Corpus Linguistics at Work*. Amsterdam; Philadelphia [Pa.: J. Benjamins.