
Le corpus lexicographique dans les langues à tradition orale: le cas du dialecte fang-mekè*

Yolande Nzang-Bié, *Département des Sciences du Langage, et Groupe de Recherches en Langues et Cultures Orales (GRELACO), Université Omar Bongo, Libreville, Gabon (yolnzang@yahoo.fr)*

Résumé: Les corpus sont à la base de la plupart des recherches en linguistique et particulièrement lexicographique. La compilation d'un corpus est une activité spécialisée dont dépend le résultat de la recherche en question. Le sujet de cet article est la compilation du corpus lexicographique dans les langues à tradition orale, et exige une démarche différente de celle ayant une longue tradition écrite. De ce fait, ces dernières disposent d'une importante documentation pouvant servir comme base pour de nombreux sujets de recherche. L'auteur propose comme approche une analyse qui permettrait de mieux rendre compte des spécificités lexicales et sémantiques des langues à tradition orale.

Par le truchement de la production orale libre, l'auteur base ses hypothèses de recherche sur une expérience en dialecte fang-mekè, une variante linguistique localisée au Gabon. Les résultats permettent de mettre l'accent sur deux données essentielles du processus de compilation dans les langues à tradition orale: les informateurs et la représentativité du corpus. Cette dernière, qui doit s'exprimer à travers des champs lexicaux diversifiés mais également équilibrés, permettrait d'élaborer des dictionnaires dans lesquels les locuteurs, qui en sont les premiers utilisateurs, doivent se reconnaître.

Mots-clés: CORPUS, LEXICOGRAPHIE, LANGUES À TRADITION ORALE, LANGUES À TRADITION ÉCRITE, INFORMATEURS, EXHAUSTIVITÉ, REPRÉSENTATIVITÉ, CHAMPS LEXICAUX, ORALITÉ, ÉCRITURE, MÉTHODE, DIALECTE FANG-MEKÈ, CORPUS ÉQUILIBRÉ.

Abstract: The Lexicographic Corpus in Languages with an Oral Tradition: The Case of the Dialect Fang-Mekè. Corpora form the basis of most linguistic and especially lexicographic research. The compilation of a corpus is a specialised activity on which depends the result of the research to be undertaken. The subject of this article is the compilation of a lexicographic corpus in languages with an oral tradition which requires a different approach from those having a long written tradition. Because of this fact, the latter have an important documentation which can serve as basis for many subjects of research. The author proposes an analytic

* Cet article a été réalisé pendant que l'auteur participait à un programme d'échange pour l'année académique 2000-2001 au Département d'afrikaans et de néerlandais à l'Université de Stellenbosch en Afrique du Sud. L'article doit beaucoup aux discussions et aux enseignements que l'auteur a eus avec professeur R.H. Gouws dont les points de vue dans le domaine lexicographique lui ont été bénéfiques et enrichissants.

approach which will allow better to account for lexical and semantic particularities of languages with an oral tradition.

Through the method of free oral production, the author bases her hypotheses of research on an experiment in the Fang-Mekè dialect, a language variant of Gabon. The results make it possible to stress two essential factors in the compilation process of languages with an oral tradition: the informants and the representativeness of the corpus. The latter which must encompass varied but also balanced lexical fields, would allow the planning of dictionaries in which the speakers who are the first users, can recognize themselves.

Keywords: CORPUS, LEXICOGRAPHY, LANGUAGES WITH AN ORAL TRADITION, LANGUAGES WITH A WRITTEN TRADITION, SPEAKERS, EXHAUSTIVENESS, REPRESENTATIVENESS, LEXICAL FIELDS, ORALITY, WRITING, METHOD, DIALECT FANG-MEKÈ, BALANCED CORPUS.

1. Introduction

Dans le domaine de la recherche linguistique, il existe une littérature abondante et variée sur la thématique du corpus et son importance capitale pour la plupart des recherches linguistiques. Aussi, notre travail n'a pas pour objectif de théoriser sur ce sujet où beaucoup a déjà été dit. Bien que l'on constate une évolution rapide des méthodes pour le perfectionnement de ce domaine, je pense comme Kennedy (1998) que la notion de ce que constitue un corpus valide reste un sujet à controverse.

La lexicographie, plus que les autres domaines des sciences du langage, n'échappe pas à ce principe. Ma réflexion a pour principal objectif de montrer le caractère primordial, essentiel, représentatif, qualitatif et quantitatif du corpus lexicographique, et davantage pour la spécificité du corpus lexicographique dans les langues à tradition orale où il existe, manifestement, peu de documents écrits pouvant constituer une base de données primaires pour l'élaboration des dictionnaires. En effet, de nombreuses références sur cette question sont basées sur des langues ayant une longue tradition écrite et bénéficiant de ce fait d'un fond documentaire divers, pouvant servir de référence à la recherche dans différents domaines de la linguistique, et particulièrement à la compilation des corpus lexicographiques. Il me semble que les langues à tradition orale, qui sont pour le grand nombre des langues africaines, devraient être abordées selon une approche relativement différente de celles à tradition écrite. Ma contribution se veut plus pratique que théorique. Elle me permet de partager mon expérience du corpus lexicographique dans les langues à tradition orale, à travers le dialecte fang-mekè, une variante linguistique localisée au Gabon, avec d'autres chercheurs que cette question intéresse. Toutefois, avant d'examiner les données du corpus lexicographique dans les langues à tradition orale, il est nécessaire de rappeler quelques principes inhérents au corpus linguistique en général et lexicographique en particulier. Aussi, ma réflexion va s'articuler sur les points suivants:

- le corpus lexicographique et ses objectifs;
- le corpus dans les langues à tradition orale;
- l'expérience du dialecte mekè; et
- la planification rationnelle du corpus lexicographique dans les langues à tradition orale.

2. Le corpus lexicographique et ses objectifs

En linguistique, de manière générale, le corpus est présenté comme la première étape dans la plupart des projets de recherche. Il constitue la source essentielle d'informations. De nombreux chercheurs le présente comme un ensemble de textes écrits ou transcrits qui sert de point de départ à la plupart des analyses ou descriptions dans une langue donnée. Cette approche est aussi celle de Dubois Charlier (1997: 312), pour qui le corpus est "une base de données textuelles réunissant des textes écrits et oraux de provenances variées et de registres divers: livres, journaux, débats etc.", ou encore celle de Francis (1992: 17) "a collection of texts assumed to be representative of a given language, dialect, or other subset of a language, to be used for linguistic analysis". Il constitue pour Kennedy (1998: 4) "an empirical basis not only for identifying the elements and structural patterns which make up the systems we use in a language, but also for mapping out our use of these systems".

En principe, le corpus a une fonction "représentative". Il doit par conséquent être représentatif du répertoire oral, grammatical et écrit maîtrisé par les locuteurs natifs performants dans leurs divers usages de la langue, et permettre de rendre compte de la totalité des caractéristiques d'une langue donnée. À cet effet, le corpus doit être l'objectif d'une étude exhaustive et non sélective; c'est ce que Quirk (1992: 458) désigne par "the vital principle of total accountability". Par ailleurs, le corpus doit aller au-delà des missions ou buts qui lui sont initialement assignés, comme le dit justement Svartvik (1992: 9): "corpora are not simply language samples that may provide useful illustrative examples but a theoretical resource."

En lexicographie, le corpus ne déroge pas aux principes observés plus haut, mais il tient sa particularité du caractère de l'exhaustivité, ce qu'on n'observe pas, par exemple, dans la description linguistique. Cette dernière utilise des corpus questionnaires¹ qui remplissent parfaitement les fonctions qui leurs sont assignées, même si l'on relève pour des descriptions plus exhaustives des insuffisances. Celles-ci sont supplées par diverses productions orales ou écrites libres qui ont pour but de vérifier la fonctionnalité des règles élaborées en contexte naturel et variable.

La lexicographie forme la structure de base d'une langue. Par ce fait, les données qui proviennent de la langue doivent refléter une représentativité et une exhaustivité sans faille, et l'on insistera jamais assez sur ce principe de

représentativité et d'exhaustivité qui caractérise le corpus lexicographique. C'est pour cette raison qu'un tel corpus doit prendre en compte les différents domaines d'application dont fait usage la langue dans ses multiples actes de parole, le résultat étant tributaire des données compilées. Cette représentativité du corpus permet de découvrir continuellement comment les langues se comportent dans toutes leurs diversités. La phase de compilation du corpus qui est l'acquisition des données matérielles pour tout projet de dictionnaire, doit être, comme l'écrit Gouws (2001) "a highly skilled activity". Le corpus lexicographique doit être en mesure de fournir en théorie tous les emplois possibles avec leur fréquence, leur dispersion, leur contexte (Dubois Charlier 1997: 312).

Il est manifeste que le corpus n'est pas une fin en soi, mais il est une source évidente de données dans un processus d'élaboration d'un dictionnaire. Toutefois, l'on ne doit pas perdre de vue que la compilation d'un corpus dépend du type de dictionnaire que l'on veut élaborer. À mon avis, tout projet de compilation d'un corpus lexicographique doit pouvoir répondre à quelques interrogations, comme

- à qui est adressé le projet du dictionnaire;
- quel type de dictionnaire veut-on élaborer; et
- quels sont les objectifs visés.

En effet, il me semble erroné qu'un projet de dictionnaire axé sur la thématique des techniques de chasse, par exemple, s'étende démesurément sur des domaines de la langue qui n'ont rien à voir de près ou de loin avec le sujet en question. Aussi, le terme de la "représentativité" doit être clairement cerné, et ne doit pas prêter à confusion. Je suis d'avis qu'un corpus, quel que soit le but visé, ne doit pas être sélectif dans le sens d'être limité à un simple questionnaire. C'est pour cette raison que j'adhère à l'approche qui stipule de "faire des provisions" dans la planification de tout corpus lexicographique. J'entends par "faire des provisions", le principe de la collecte exhaustive des données qui permettrait, par la suite, d'extraire du corpus les données dont on a besoin pour la thématique concernée. Cette approche que je tiens du professeur R.H. Gouws présente l'avantage que le corpus puisse être utilisé pour d'autres sujets. C'est dans ce contexte que l'on peut parler du caractère multifonctionnel d'un corpus.

3. Le corpus lexicographique dans les langues à tradition orale

De l'avis de nombreux chercheurs, le corpus lexicographique doit s'orienter sur deux axes parallèles: les sources écrites et les sources orales. Le constat que l'on est amené à faire, c'est qu'une place prioritaire est accordée aux sources écrites. Cette approche est celle qu'on retrouve dans les propos de Kennedy (1998: 7): "In this case of corpus based research, the evidence is derived directly from

texts"; mais aussi d'autres sources tel que le Bureau du WAT (1999: 4), qui écrit: "The most common source of language material is the written text. Because words are available in their full context, the texts are readily accessible and can be utilized immediately. Written sources include the following: published sources such as books, journals and magazines, newspapers, reports and proceedings; and unpublished sources (ephemera) such as advertising material, financial reports, pamphlets, minutes and letters. Included in the published written sources are other dictionaries and word lists (technical as well as general) which the editor may consult during the compilation process and which may also serve to check whether all important words and terms have been included"; et plus loin encore, on peut lire: "it is advisable to start a material collection with written materials as it is more easily collected and processed". Cette approche qui s'applique parfaitement aux langues ayant une longue tradition écrite, ne s'adapte pas, à mon avis, aux langues à tradition orale pour lesquelles l'essentiel de la communication dans ses divers usages est centré sur l'oralité et où, dans la plupart des cas, il n'existe pas ou très peu de données écrites. Bon nombre de langues du Gabon pour lesquelles il n'existe aucune documentation écrite peuvent l'illustrer et dans les cas où il y aurait des traces, celles-ci sont limitées à quelques fascicules dont la graphie est souvent douteuse et ne reflète pas la situation de la prononciation réelle. Il me semble que les sources écrites dont disposent actuellement certaines langues à tradition orale, sans pour autant être négligées, doivent faire l'objet de vérifications par de nouvelles enquêtes, en vue de leur actualisation, afin d'aboutir à un travail qualitatif et référentiel.

Une autre catégorie de sources orales vers laquelle l'on devrait également s'orienter, sont de nombreuses bandes sonores entreposées dans des musées et les institutions de communication (radios, télévisions). Celles-ci sont de nombreuses sources d'informations dans des domaines divers, mais elles sont généralement inexploitées par manque de transcriptions. Ce problème de transcription du matériel oral disponible mais non-exploité, représente l'une des difficultés que l'on rencontre dans la compilation du corpus dans les langues à tradition orale.

Aussi, fort de ce qui précède, j'encourage fortement pour les langues à tradition orale l'approche inverse plus adaptée et rationnelle qui considérerait comme

- les sources primaires, les données orales (qu'elles soient à compiler ou à transcrire); et
- les sources secondaires, les données écrites.

Dans la première catégorie de données, la langue est fonctionnelle, dynamique et pleinement efficace. Cette approche permet de saisir la langue dans son contexte naturel et de capter les domaines d'usage qui constituent l'activité culturelle du peuple, ce qui est l'objet de notre étude. Dans cette activité de compi-

lation du corpus, il est question de la compétence aussi bien que de la performance des locuteurs. Les sources écrites, qui sont également un aspect capital, doivent suppléer les sources orales. Elles contiennent indéniablement de nombreuses informations. C'est dans ce contexte bi-directionnel que les principes de représentativité, de diversité et d'exhaustivité tiennent toute leur importance.

L'approche du questionnaire comme élément de base pour la compilation d'un corpus lexicographique a perdu son importance primordiale. Limité par ce fait, un questionnaire, aussi exhaustif soit-il, ne peut prendre en compte l'essentiel des données d'un domaine spécifique, tout en restant subjectif du fait qu'un questionnaire représentatif pour un chercheur, ne l'est pas forcément pour un autre. Toutefois, je pense que le "corpus questionnaire" peut servir de point de départ dans le contexte d'un corpus lexicographique thématique ou dans un domaine spécialisé qui exige une terminologie spécifique. Mais il doit, nécessairement être renforcé par ce que j'ai appelé "les données orales à production libre". Elles présentent l'avantage d'être produites de manière spontanée et dans leur contexte naturel. Le questionnaire peut également servir de point d'appui dans la recherche des mots dont les définitions ou les portées sémantiques ne figurent pas explicitement dans le corpus global.

De tout temps, la linguistique a fonctionné à l'aide de "corpus questionnaire". Si cette approche reste encore valable pour certaines disciplines de la linguistique, et de la lexicographie des langues à tradition orale, la compilation des données est un processus dynamique et continu. Toutefois, elle doit prendre en compte l'inquiétude formulée par le WAT (1999: 6), qui est aussi mienne: "be aware of an overabundance of language material or inadequate filtering methods, which may slow down the process of manuscript making."

4. L'expérience dans le dialecte fang-mekè

Mon expérience sur le corpus en dialecte fang-mekè est à l'origine de cette réflexion sur le corpus lexicographique dans les langues à tradition orale. Elle a été motivée par une formation en lexicographie à l'Université de Stellenbosch en Afrique du Sud. C'est en prévision de cette formation que j'ai procédé à une enquête libre auprès d'un groupe de locuteurs natifs fang, du dialecte mekè, sans trop lui donner un objectif précis.

Pour la localisation du groupe-cible, l'on retiendra brièvement que le dialecte mekè est l'une des composantes du groupe linguistique fang². Il est localisé, selon diverses sources (entre autres Mba-Nkoghe 1981, Voltz 1990, Ondo-Mebiame 1992) à Libreville et Cocobeach dans la province de l'Estuaire, mais aussi à Mitzic dans la province du Woleu-Ntem. Démographiquement, je ne dispose pas de données fiables sur ce groupe de locuteurs.

Sur le plan documentaire, le dialecte mekè présente très peu de documents écrits. Actuellement, hormis quelques descriptions linguistiques non exhaustives et des fascicules laissés par des missionnaires, le document le plus impor-

tant écrit en dialecte mekè est la traduction du nouveau testament biblique (1962). Cet ouvrage contient également des cantiques, des prières et des oraisons en dialecte mekè. Cette source écrite est d'une grande valeur quand on sait que la Bible est considérée depuis le 18^e siècle comme l'une des premières pièces significatives dans la recherche des corpus (Kennedy 1998).

L'expérience menée sur le dialecte mekè a été faite à base d'un matériel très simple, à savoir un enregistreur et des cassettes audio, ce qui est une des techniques recommandées (Kennedy 1998: 20).

4.1 La méthode utilisée

J'ai procédé à une enquête à "production libre" auprès d'un groupe de locuteurs natifs performants, ayant reçu leur éducation dans la langue maternelle et n'utilisant le français qu'en cas de besoin spécifique. Les entretiens se sont déroulés de manière libre et collégiale entre les différents locuteurs. Chacun était libre de développer sa pensée, mais il pouvait à tout moment être interrompu par un autre locuteur pour compléter l'information, expliciter les propos du prédécesseur et même les rectifier, ajouter ou introduire de nouvelles données. Cette méthode a fait que j'ai assisté à un échange actif et fructueux. L'un des constats que j'ai eu à faire est la spontanéité dans les réactions des différents locuteurs.

Après avoir expliqué à mes interlocuteurs quelles étaient mes attentes, j'ai introduit mon enquête sur la thématique de l'histoire des Fang. À la suite de ce thème, d'autres se sont spontanément développés sans que cela ait été planifié et ce dans une logique déconcertante. C'est ainsi que l'on pouvait passer, par exemple, du thème du mariage à celui de la mort ou encore de la nature, sans qu'il y ait rupture dans le développement des idées chez mes différents interlocuteurs. Tout se présentait comme si tous les thèmes abordés "inconsciemment" étaient le processus logique d'une même chaîne et dont on passait, sans rupture, d'un maillon à un autre.

Tout au long des échanges, j'ai mené les débats de sorte que les différents interlocuteurs mais aussi l'enquêteur pouvaient, sans dérouter l'intervenant, poser des questions pour d'amples informations. En exemple, les locuteurs hommes utilisaient énormément de proverbes pour illustrer leurs propos. Aussi, ai-je à un moment interrompu mon locuteur pour avoir des explications sur le proverbe. J'ai eu droit à une longue définition du proverbe qui a été suivi par une série d'autres proverbes, énoncés et explicités par mes locuteurs. De fait, les explications sur un point pouvaient conduire à un autre thème et les interruptions servaient à faire des remarques sur des points précis.

À propos de ces interruptions, j'ai relevé la fréquence des expressions telles que: "je peux ajouter à ce qu'il vient de dire que ...; c'est aussi ce qu'on appelle en fang³; ce que tu dis n'est pas complet ...; tu as omis de préciser que ...; ton explication n'est pas juste ...; tel type de proverbe peut aussi illustrer ton propos ...; ce mot n'est pas de notre dialecte nous l'avons pris chez ...; est ce que quelqu'un à quelque chose à ajouter ...; si je me trompe que untel ou untel me

reprenne ...; je ne pense pas tout connaître sur le sujet, mais je peux dire que ..." et bien d'autres encore.

J'ai noté que mes locuteurs n'étaient pas omniscients. Dans certains domaines spécifiques, les informateurs me renvoyaient vers quelqu'un de plus compétent avec des expressions telles que "je ne connais pas grand chose dans ce domaine, mais tu peux aller voir untel qui peut très bien t'en parler; cependant je peux dire que ...", ce qui me laisse supposer que tout locuteur compétent et performant n'est jamais totalement ignorant des concepts relatifs à son milieu.

Il est significatif de souligner que pendant les entretiens, les locuteurs pouvaient se lever librement, vaquer à quelques occupations pas très prenantes et revenir rejoindre le groupe sans qu'il y ait rupture dans le déroulement des entretiens. Cette méthode présente ainsi l'avantage de ne pas être contraignante, aussi bien pour les enquêteurs que pour les informateurs.

À la fin du premier entretien, j'ai constaté que mes informateurs, réticents au départ, avaient été très heureux de me fournir ces informations. Ils donnaient l'impression d'avoir contribué à quelque chose de constructif et pérenne pour leur langue. Aussi, étaient-ils entièrement disposés pour la suite de l'enquête.

L'expérience que je viens de relater n'est certainement pas celle d'autres chercheurs qui se sont intéressés à cette problématique, mais je pense que si elle est planifiée, cette méthode de travail dans la compilation du corpus lexicographique dans les langues à tradition orale, devrait mener à des résultats probants.

4.2 Les résultats

Pour une période assez courte de quatre jours, j'ai pu enregistrer quatre cassettes audio d'une durée de 1h30mn chacune. L'objectif de ces enregistrements n'est pas de faire le traitement lexicographique⁴ en vue d'établir des listes de fréquence, de concordance, de fréquence d'un niveau de langue, moins encore l'élaboration du dictionnaire. Il vise essentiellement les divers champs lexicaux ou domaines de recherche que ma méthode et mon expérience m'ont permis de mettre en évidence, et qui constituent, à mon avis, un échantillon des données lexicales pour beaucoup de langues africaines à tradition orale.

Les champs lexicaux qui ont pu être extraits du corpus sont représentés dans le tableau 1.

Tableau 1: Champs lexicaux et sous-thèmes du corpus

Thèmes représentant des termes superordonnés	Hyponymes ou sous-thèmes
<ul style="list-style-type: none"> -- histoire -- les récits⁵ 	contes, proverbes, récits épiques, devinettes

<ul style="list-style-type: none"> -- le mariage 	<p>le choix du conjoint, les présentations, les fiançailles, la dot, mariage par rapt, mariage par consensus</p>
<ul style="list-style-type: none"> -- éducations des enfants (jeunes filles et garçons) -- la palabre -- la mort 	<p>l'annonce, l'enterrement, le deuil, la levée de terre, le port du deuil, le retrait de deuil, les funérailles</p>
<ul style="list-style-type: none"> -- la procréation 	<p>la grossesse, l'accouchement, les matrones, le bain rituel</p>
<ul style="list-style-type: none"> -- l'adultère -- le veuvage -- la réinsertion sociale -- la sémantique des noms propres -- le maquillage (extrait les termes relatifs à la beauté) 	
<ul style="list-style-type: none"> -- le décompte du temps 	<p>heures, jours, semaines, mois, saisons, années</p>
<ul style="list-style-type: none"> -- la destinée -- la circoncision -- les travaux champêtres -- rituels initiatiques -- la pêche 	<p>les techniques, les instruments, la pêche par les hommes, la pêche par les femmes, les périodes</p>
<ul style="list-style-type: none"> -- la chasse 	<p>les techniques, les instruments, les types, les périodes, les techniques de pièges</p>
<ul style="list-style-type: none"> -- les constructions -- l'occupation de l'espace -- les astres -- l'artisanat -- l'éducation sexuelle -- le corps humain -- les aliments -- l'écosystème -- les couleurs 	<p>les techniques, les matériaux, les types</p>

J'ai noté, et cela est peut-être une particularité de mon groupe d'informateurs, qu'ils évitaient subtilement de parler des questions relatives aux sexes ou aux organes génitaux⁶, même lorsque les échanges devaient les y conduire naturellement. J'ai personnellement évité d'insister sur ce point pour ne pas heurter la sensibilité des informateurs. Je pense que si la tendance se confirme dans d'autres expériences, cet aspect, dans l'enquête, peut faire l'objet d'un traitement particulier à définir par les enquêteurs.

Cet échantillon de corpus qui n'est ni exhaustif, ni représentatif est, à mon avis, une ébauche, une esquisse de ce qui pourrait constituer un corpus lexicographique des langues à tradition orale. Il a pu faire ressortir quelques thèmes saillants, ce qui dénote que menée sur une échelle représentative, cette méthode peut conduire à une vue globale et non moins précise des champs lexicaux récurrents et fonctionnels permettant d'aboutir à un corpus équilibré.

5. Pour une planification du corpus lexicographique rationnelle

L'activité de compilation d'un corpus lexicographique, dans les langues à tradition orale, doit prendre en compte deux données essentielles: les informateurs et la représentativité du corpus équilibré.

5.1 Les informateurs

Les informateurs locuteurs natifs constituent le pilier central pour toute compilation d'un corpus lexicographique. Dans le cadre d'un corpus lexicographique, le nombre d'informateurs doit varier en fonction de l'ampleur du projet à mener. Dans tous les cas, de par son ampleur et sa diversité, aucun corpus lexicographique, aussi restreint soit-il, ne peut être compilé par un seul informateur ou par un seul groupe d'informateurs. Le nombre doit aller croissant selon les besoins du projet mené. En effet, l'univers linguistique est extraordinairement vaste et complexe, et aucun humain n'a la prétention de le maîtriser. À mon avis, et d'accord avec Chafe (1992: 80) "it is impossible for any individual to understand more than a tiny segment of it. Different individuals understand different tiny segments — some perhaps slightly more than others — but no human understanding can go very far. Understanding may be enlarged when it is distributed over groups of individuals". Cette approche de Chafe plaide en faveur de la diversité dans le choix des informateurs.

Le dictionnaire reflète la compétence, la dynamique linguistique d'une société donnée. Il véhicule l'activité culturelle d'un peuple et doit, par conséquent, refléter les différents domaines d'activité dont fait usage une société donnée. En effet, toute élaboration d'un dictionnaire doit être adaptée aux besoins de la société qui en est le premier utilisateur. Celle-ci doit se reconnaître dans le dictionnaire constitué à son usage. À ce titre, le corpus, à travers ses informateurs, doit être représentatif de la langue dans ses différents domaines comme un tout. Toutefois, l'on doit garder en esprit qu'aussi exhaustives qu'elles soient, les données d'un corpus ne seront jamais qu'un minuscule échantillon de toute l'activité linguistique reçue ou produite par les locuteurs d'une langue donnée. Ce qui est essentiel, à mon avis, c'est que cet échantillon soit représentatif de l'activité humaine à travers ses locuteurs. Je partage à cet effet la réflexion de Sinclair (1991: 20) qui écrit que "corpora of finite size were inherently deficient because any corpus is such a tiny sample of language".

Les informateurs performants locuteurs natifs de la langue vers laquelle le

projet est mené doivent être majoritairement des notables, des personnes d'âges mûrs, des spécialistes de domaines spécifiques des deux sexes usant le moins possible une langue importée, à l'instar du français; car comme nous l'avons déjà dit, l'activité de compilation d'un corpus est une question de performance, mais aussi de compétence.

Les informateurs doivent provenir principalement de deux milieux: le milieu rural et le milieu urbain. Cette répartition, source d'enrichissement pour le corpus, est importante, car elle permet de prendre en compte les différents milieux naturels dans lesquels vivent quotidiennement les informateurs, et qui sont susceptibles d'influencer, d'une manière ou d'une autre, leurs performances linguistiques. Aussi, le choix de locuteurs dans un univers limité peut fausser la représentativité du corpus et déséquilibrer les données.

Les informateurs doivent travailler dans un cadre naturel, non contraignant et le plus souvent mis en situation de conversations, d'entretiens, à travers lesquels les différents locuteurs s'expriment librement. L'une des approches que je propose est la participation des enquêteurs à des cérémonies ou des rencontres en milieu et en contexte naturels. Mais on peut aussi organiser, comme c'est généralement le cas, des séances "artificielles"⁷ de rencontres. On devrait éviter, le plus possible, de faire débiter sans motivation des listes de mots aux informateurs, ce qui par expérience finit par lasser ceux-ci qui peuvent aller jusqu'à demander l'arrêt de l'enquête. Toutefois l'on doit veiller à éviter des situations cacophoniques où l'on ne sait plus qui dit quoi, de même qu'on doit veiller à créer un climat de confiance d'une part entre les enquêteurs et les informateurs, et d'autre part entre les différents informateurs. Cette atmosphère est nécessaire au bon déroulement de la compilation du corpus.

5.2 La représentativité du corpus équilibré

Ce travail n'a pas la prétention d'aborder la grande question du contenu du corpus lexicographique en ce qui concerne le nombre de mots, les différents types de mots que doit contenir un corpus ou encore le nombre de mots d'un exemple. De nombreuses références disponibles en font état, et il ressort qu'un bon corpus se chiffre en dizaines de millions de mots (Kennedy 1998). Je retiens, par exemple, que Sinclair (1991: 20) a suggéré que "10–20 million words might constitute a useful small general corpus".

Ce travail se propose de mettre l'accent sur la représentativité du corpus lexicographique, en termes de diversité de champs lexicaux dans les langues à tradition orale.

Les champs lexicaux sur lesquelles doivent se baser une compilation de corpus dans une société donnée ne sont pas facilement décelables. Cette tâche est très difficile quand on sait que, même sous le contrôle de toute une communauté, il est difficile, voire impossible de recenser les différents domaines d'emploi du langage. Kennedy (1998: 72) a proposé le tableau ci-après pour les catégories de textes oraux ("categories of spoken texts").

Tableau 2: Catégorie des textes oraux (selon Kennedy 1998)

— Monologue
— formal
- written to be read aloud
- prepared but unscripted public speech
— less formal
- academic lectures
- commentaries on public occasions
- sports commentaries
- demonstrations
— Dialogue
— face-to-face dialogue
— public discussion (e.g. on radio)
— business or professional transactions, e.g.
- client and professional
- workplace interaction between colleagues
- commercial (sales) transaction
— informal
- within family
- between friends
— telephone dialogue
- between interlocutors known to each other
- between interlocutors unknown to each other
— structured interaction
- interview
formal, e.g. arts programme
less formal, e.g. with survivors and witnesses to events such as accidents
- debate
- committee meeting

Cette liste permet de voir que bien de sujets sont facilement applicables aux langues à tradition orale. Toutefois à cette liste, j'ajouterai les sujets suivants contenus dans le tableau 3.

Tableau 3: Catégories des données orales (en complément du tableau 2, selon l'auteur)

— Monologue
— émission radio ou télévisée
- présentation des nouvelles, spot publicitaires, annonces publiques, annonces culturelles, jeux éducatifs
- musique
- informations documentaires musique
- films ⁸
- théâtre ⁹
— Dialogue
— production libre (permet d'accéder à divers thèmes) ¹⁰

- participation à diverses cérémonies
 - rituelles
 - funéraires (deuil, levée de terre, veuvage, retrait de deuil, etc.)
 - folkloriques
 - initiatiques
 - religieuses
 - règlement de palabre
 - tribunal traditionnel
 - maquillage (termes relatifs à la beauté)
 - soins de santé
 - purification
 - divination, voyance
 - mariage
 - fêtes
 - pêche, chasse, récolte, agriculture
 - naissances
 - installation de notables
 - transactions de types divers
 - politiques
 - économiques
 - culturelles (production de différents genres littéraires, jeux collectifs)
 - détente

Cette liste a une fonction purement indicative et référentielle, car comme l'écrit Kennedy (1998: 72) "taken over a whole community, it is simply impossible to get a balance which represent actual use of the language". Les dictionnaires de langues africaines doivent refléter des réalités africaines. Aussi, cette liste fournit des pistes de recherches pouvant aider dans la compilation des thèmes relatifs au corpus lexicographique, en mettant l'accent sur des facettes courantes des langues à tradition orale.

J'ai insisté, tout au long de ce travail, sur le défi de la représentativité du corpus lexicographique. Un autre défi est celui de la notion de corpus équilibré. Cette notion n'est pas considérée dans le texte, tenant compte des nombreuses sources, et la proportion des données équitables entre les sources écrites et les sources orales ("60% spoken texts and 40% written texts", Kennedy 1998: 62). Elle est retenue selon la répartition équitaine des données à l'intérieur d'un même sujet qu'il soit spécifique ou général. Le but du corpus est d'être fonctionnel. Aussi, un projet de compilation d'un dictionnaire général doit prendre en compte, non seulement tous les domaines possibles repérables dans la langue, mais aussi des aspects identifiables à l'intérieur de chaque domaine. Lorsque le projet est spécifique, les différents aspects, même ceux qui paraissent anodins, doivent être exploités. Cela permet d'aboutir à un corpus fiable.

6. Conclusion

J'ai examiné dans ce travail la problématique du corpus dans les langues à tradition orale pour lesquelles l'essentiel de la communication est basé sur la pro-

duction orale. Je reste consciente que ce type d'enquête est dur et coûteux, du fait qu'il nécessite des moyens matériels, financiers et surtout d'énormes ressources humaines.

L'activité de l'élaboration d'un corpus lexicographique n'est pas une activité nouvelle, mais il semble aujourd'hui encore que les langues à tradition orale n'ont pas réellement trouvé leur place dans les banques de données existant sur ce sujet. Aussi, dans ce siècle de mondialisation, il est plus qu'imprévisible que les lexicographes et particulièrement les lexicographes africains donnent aux langues à tradition orale une place référentielle dans le traitement du langage par le truchement de la compilation des corpus lexicographiques. Pour que cette tâche soit effective, je pense comme Gouws (2001) que les unités lexicographiques doivent prendre les dispositions suffisantes dans leur planification pour cet aspect important de leur entreprise lexicographique. Ceci ne revient pas à jeter un discrédit sur ce qui a déjà été fait, mais d'améliorer les données disponibles afin de constituer des bases de données lexicales et thématiques considérables pour les langues à tradition orale. Il ne s'agit pas de "réinventer la roue", mais d'attirer l'attention de nombreux lexicographes africains sur une problématique réelle. Botha (2000) parle d'une coopération franche entre les lexicographes, mais aussi entre les lexicographes et les utilisateurs des dictionnaires afin d'éliminer les obstacles entre les utilisateurs et les concepteurs. J'ai proposé des méthodes de recherche qui peuvent amener à mieux aborder ce problème. L'accent a été mis sur des points prioritaires dans la compilation du corpus dans les langues à tradition orale.

Ces quelques pages ont posé une réflexion dans la perspective du développement rationnel et fonctionnel des langues à tradition orale par l'usage de cet outil indispensable qu'est le dictionnaire. Je pense que ces langues ont besoin de compiler, pour l'instant, des données fonctionnelles menant à l'élaboration des dictionnaires, mais permettant de constituer des bases de données qui aideraient à produire, à long terme, des travaux aussi importants que les dictionnaires monolingues, des encyclopédies et bien d'autres ouvrages de référence dans des domaines divers. Aussi, cette réflexion de Botha (2000) pourrait trouver toute sa signification: "the much talked about African Renaissance will, through lexicographic co-operation, be characterised by emergence of dictionaries for all African languages."

Notes

1. Je désigne par corpus questionnaire, des listes de questions établies pour l'élaboration des systèmes phonologiques, morphologiques et syntaxiques. L'on parle ainsi du *Questionnaire d'enquête linguistique* de J. Greenberg et W.E. Welmers (traduit par Jean Doneux), Université de Dakar, 1967.
2. Le groupe linguistique fang est classé chez Guthrie (1948), où il porte le sigle A75. Il est constitué, en plus du mekè, de cinq autres dialectes que sont: l'atsi, le dzaman, le mveny, le ntumu et l'okak.

3. Mes locuteurs parlent invariablement le fang (qui est la désignation du groupe linguistique) ou le parler mekè (nkobe wa meke (e)ne: notre parler meke d'ici).
4. Je consacre cet autre aspect à une recherche qui est en cours et qui permettra de faire paraître, avec les données disponibles élargies, un dictionnaire parallèle du dialecte mekè.
5. Les divers récits sont constitués de thèmes divers. J'ai relevé ceux relatifs à la jalousie, la cupidité, la haine, la méchanceté, la sorcellerie, l'amour, l'amitié, la confiance, la violence, l'autorité, le vampirisme, la palabre, la faune, la flore, les interdits, la nature, le respect, la patience, la ruse, l'adresse, la force, l'épreuve, l'initiation, les classes d'âge, les jeux, la notion du temps, la tradition.
6. Chez le peuple Fang, les interdictions relatives aux sexes sont culturelles et relativement taboues en public.
7. Je désigne par séances artificielles, des séances d'enregistrements qui ne se déroulent pas dans un contexte naturel.
8. et 9. sont classés dans cette catégorie du fait que l'enquêteur n'a pas accès aux informateurs.
10. Un aperçu de cet aspect à été donné à la section 4.

Bibliographie

- Atkins, B.T.S. 1994. A Corpus-based Dictionary. Grundy, Marie-Hélène Corréard-Valerie (Éd.). 1994: xix-xxxii.
- Atkins, Sue et Jeremy Clear. 1992. Corpus Design Criteria. *Literary and Linguistic Computing* 7: 1-16.
- Botha, Willem. 2000. We All Stand Together, Don't We? — African Renaissance through Dictionaries. Heid, Ulrich et al. (Éds.). 2000: 163-168.
- Bureau of the WAT. 1999. Language Material Collection and the Compilation of a Data Base. Notes de cours non publiées.
- Chafe, Wallace. 1992. The Importance of Corpus Linguistics to Understanding the Nature of Language. Svartvik, Jan (Éd.). 1992: 79-96.
- Clear, Jeremy. 1996. Technical Implications of Multilingual Corpus Lexicography. *International Journal of Lexicography* 9(3): 265-273.
- Crowdy, Stephen. 1993. Spoken Corpus Design. *Literary and Linguistic Computing* 8: 259-265.
- Dubois Charlier, Françoise. 1997. Compte rendu de l'Oxford-Hachette French Dictionary. *International Journal of Lexicography* 10(4): 311-329.
- Francis, W. Nelson. 1992. Language Corpora B.C. Svartvik, Jan (Éd.). 1992: 17-32.
- Gouws, R.H. 2001. Lexicographic Training: Approaches and Topics. Emejulu, James Duplessis (Éd.). 2001. *Éléments de lexicographie gabonaise. Tome I: 58-94*. New York: Jimacs-Hillman.
- Guthrie, Malcolm. 1948. *The Classification of the Bantu Languages*. Londres: Oxford University Press.
- Grundy, Marie-Hélène Corréard-Valerie (Éd.). 1994. *Oxford-Hachette French Dictionary*. Oxford: Hachette Oxford University Press.
- Heid, Ulrich et al. (Éds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000. Stuttgart, Germany, August 8th-12th 2000*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. Londres: Longman.
- Koček, Jan, Marie Koprivova et Vera Schmiedtova. 2000. The Czech National Corpus. Heid, Ulrich et al. (Éds.). 2000: 127-132.

- Mba-Nkoghe, J.** 1981. *Phonologie et classes nominales en fang (Langue bantoue de la zone A)*. Thèse 3^e cycle. Paris: Université Sorbonne Nouvelle.
- McEnery, Tony et Andrew Wilson.** 1996. *Corpus Linguistics*. Édimbourg: Edinburg University Press.
- Ondo-Mebiame, P.** 1992. *De la phonologie à la morphologie du fang parlé à Aboumezok (langue Bantu A. 78)*. 2 volumes. Thèse de doctorat. Tervuren: Université Libre de Bruxelles.
- Quirk, Randolph.** 1992. On Corpus Principles and Design. Svartvik, Jan (Éd.). 1992: 463-469.
- Svartvik, Jan.** 1992. Corpus Linguistics Comes of Age. Svartvik, Jan (Éd.). 1992: 7-13.
- Svartvik, Jan (Éd.).** 1992: *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82*. Stockholm: Mouton de Gruyter.
- Sinclair, John.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Voltz, M.** 1990. Fanj (Ntumu). *Revue Gabonaise des Sciences de l'Homme* 2: 143-173.
- Zgusta, Ladislav.** 1971. *Manual of Lexicography*. La Haye: Mouton.