

Construction des corpus bilingues (langues africaines - français) : étude contrastée des interférences phonético-sémantiques

## **Introduction**

Le français et l'anglais sont les langues officielles du Cameroun et appartiennent à la grande famille Khoisane et leurs statuts dans les villes plurilingues varient de langues maternelles à langues véhiculaires. Par contre, les 252 langues camerounaises (atlas linguistique du Cameroun, 2012) appartiennent à trois des quatre grandes familles de langues. Les langues nationales de cette étude appartiennent au phylum niger-Kordofan et sont des langues bantoues dotés de consonnes complexes et de tons à valeurs lexicales et grammaticales. Elles influencent à priori l'articulation et la sémantique du français lorsque celle-ci est langue seconde. Toutefois, les études sur les interférences se basent souvent sur des études sociolinguistiques et celles de corpus.

Comme Branca-Rosoff, Fleury, Lefevre et Pires (2011) le rappellent, la linguistique de corpus est toujours une construction qui choisit et organise les matériaux. Pour construire ces matériaux, cet article ambitionne de construire un corpus bilingue et d'en déduire les opportunités d'analyse interférentielles.

### **1. Problème**

La majorité des recherches sur le lexique en général bénéficie d'un large accès aussi bien en ligne qu'en manuscrit. Toutefois, si les langues de grande diffusion bénéficient d'une visibilité sur le web 2.0, les langues africaines peinent à généraliser la collecte des données en ligne. Les raisons émises sont de plusieurs ordres surtout en matière de langues bantoues dont les tons influencent fortement l'accent en français tout en provoquant des confusions de sens. L'une des raisons repose sur la difficulté à transcrire celles-ci et surtout à conserver les mêmes codes lors des échanges en ligne.

Il convient donc de s'interroger sur la manière de sélectionner des informateurs en ligne afin de construire un corpus bilingue servant à une analyse contrastée des morphèmes articulés.

### **2. Méthode de construction du corpus**

La sélection des informateurs a pris son ancrage sur les échanges des jeunes lycéens sur les réseaux sociaux. Pour cela, plusieurs activités ont été proposées pour obtenir ces corpus : entretiens semi-dirigés auprès des lycéens habitant les quartiers périphériques de la ville de Yaoundé et présentant un ancrage profond dans leurs langues maternelles locales, des exercices de description orale et d'échanges oraux avec leurs camarades.

#### **a) Les informateurs**

La centaine d'informateurs retenus a permis de caractériser s'ils étaient des locuteurs natifs ou non des langues camerounaises suivantes : ewondo, mengisa, basaa. En marge, s'ils ne l'étaient pas, la sélection reposait sur le nombre d'années séjournés dans le quartier et à Yaoundé. L'âge

variait entre 12 et 16 ans pour ne conserver que les pré-adolescents et les adolescents. Ce profil sociolinguistique de base a garanti la validité des données fournies par ces derniers.

#### b) Les outils et activités d'enquête

Comme premier outil d'enquête, l'interview semi-dirigée a permis d'opter pour des activités d'expression orale libre comme la description ou la narration. Il leur était demandé de décrire leur famille, leur quartier, leurs amis proches et leurs rêves. Pour certains, l'activité consistait à raconter une histoire sur les faits culturels alentours. Ces activités étaient demandées en français et en langues maternelles, donc ces derniers effectuaient une forme de version. Toutes les interviews ne dépassaient pas 5mn de manière à ce qu'après une cinquantaine, nous soyons déjà en mesure d'identifier les interférences les plus récurrentes à l'oral et le reste des interviews confirmait et/ou complétaient les premières analyses. Ainsi, il est loisible d'observer les interférences et de convenir avec Weinreich (1953, 1) sur le problème des déviations de la norme établie par une langue déterminée, de la part des personnes bilingues, en raison de leur familiarité avec plus d'une langue.

### 3. Dépouillement

Le dépouillement a débuté avec 50 enregistrements à transcrire et à analyser en fonction des interférences phoniques et sémantiques. Le dépouillement des données orales a débuté avec le logiciel Transcriber et a été finalisé par nos soins au moment de transcrire les langues africaines (beti-fang avec l'ewondo, mengisa, basaa) avec Unicode. Il s'avère après analyse que, bien que les corpus aient été constitués, les langues africaines ont été tronquées par des mots s'approchant des langues indo-européennes. Les accents des lycéens ont entraîné des transcriptions erronées car l'accent local emprunté à la L1 trop présente pendant l'articulation des mots en français. Les corpus à terme, pourraient être disponibles et compatibles avec le logiciel Lexico.

### 4. Des interférences à partir des corpus

#### a) Interférences phoniques

Il est constaté que d'une manière générale, le [i] tend à remplacer le [u] basaa en français comme dans [sitié] au lieu de « situer », le [p] fonctionne en lieu et place de [b] comme dans [pa :masi] au lieu de « pharmacie », les [e] du français sont généralement prononcés [ɛ] et surtout l'accent chantant et lent de l'ewondo est perceptible dans les paroles des lycéens.

#### b) Interférences sémantiques

Ces interférences naissent du remplacement de certains sons par d'autres tout en entraînant des changements de sens à l'exemple de [munie] (qui signifie en possession de) prononcé [mini] (qui signifie court) par un Bassaphone.

Mots clés : corpus bilingue, échantillonnage, phonético-sémantique, interférences, corpus en ligne

## **Bibliographie**

Blanche-Benveniste C. (1996), De l'utilité du corpus linguistique. *Revue Française de Linguistique Appliquée*, I-2, 25-42.

Boum Ndong-Semengue, M.A. (2013), Le paysage linguistique du Cameroun. *Langues nationales en situation. Réflexions pour la revalorisation des langues premières*, pp. 27-36.

Bourigault D., Chodkiewicz C., Humbley J. (1999), Construction d'un lexique bilingue des droits de l'homme à partir de l'analyse automatique d'un corpus aligné. *Terminologies Nouvelles*, 19, 70-77.

Goffin, R. (2005), Quels corpus et quelles approches pour une description contrastive de l'eurolecte ? *Préactes des 7es journées scientifiques du réseau Lexicologie, Terminologie et Traduction*. AUF-LTT, Bruxelles. Cité d'après <<http://perso.univ-lyon2.fr/thoiron/>>.

Habert B., Nazarenko A., Salem A. (1997), *Les linguistiques de corpus*. Paris, Armand Colin.

Kraif, O. (2006), Qu'attendre de l'alignement de corpus multilingues ? *Traduire*, 210, 17-37.

Ndibnu Messina Ethé, J. et Elobo Onana, (2014), Des interférences aux erreurs morphosyntaxiques en Langue Seconde : quelques stratégies de correction en contexte plurilingue, *Synergies Afrique des Grands Lacs*, n°3, pp. 179-192.

Pearson, J. (2000), Une tentative d'exploitation bi-directionnelle d'un corpus bilingue. *Cahiers de grammaire*, 25, 53-69.

Rastier, F. (2005), Enjeux épistémologiques de la linguistique de corpus. *La linguistique de corpus*, Rennes, Presses Universitaires de Rennes, 31-45.