

**LA FRANCOPHONIE MULTIVARIÉE  
OU  
COMMENT MESURER LES FRANÇAIS EN AFRIQUE ?**

**Sascha Diwersy**  
Université de Cologne

### **1. Introduction**

« Mesurer les français en Afrique », voilà un projet bien trop ambitieux pour pouvoir être abordé de façon adéquate dans un article de taille forcément réduite. D'ambition plus modeste, cette contribution se propose de présenter quelques éléments méthodologiques issus d'un programme de recherches qui vise à explorer la variation du français écrit dans les pays de la francophonie africaine, tout en s'inspirant d'une approche dite « à dominante inductive » (de l'anglais *corpus driven approach*).

Le présent article se divise en trois parties principales.

La section 2 donnera un aperçu des visées thématiques, méthodologiques et théoriques du projet AFRIVAR (« Variation et variétés du français en Afrique »), mené actuellement par une équipe du Centre de recherche interdisciplinaire sur la France et la Francophonie (CIFRA) de l'Université de Cologne (en collaboration avec le LDI, Paris-XIII).

La section 3 traitera ensuite du dispositif méthodique développé dans le cadre du projet, en présentant la base de corpus numérisés disponibles ainsi que les procédés et applications lexicométriques mis en place en vue de son exploitation.

La section 4 exposera les premiers résultats d'une étude de cas portant sur la combinatoire lexico-syntaxique d'un groupe de noms abstraits sous l'aspect des convergences et divergences observables à travers un ensemble d'une dizaine d'échantillons différenciés par origine géographique. Une brève conclusion clora ces observations.

### **2. Le projet AFRIVAR – visées thématiques, méthodologiques et théoriques**

Dans son ensemble, le projet AFRIVAR, initié il y a un an par le laboratoire CIFRA de l'Université de Cologne, s'inscrit dans le cadre de deux courants de recherches. L'un, évidemment bien connu en France, relève des travaux menés depuis les années 80 par S. Lafage, A. Queffélec et d'autres chercheurs associés qui ont abouti à toute une série de contributions détaillées concernant la situation sociolinguistique et les particularités (notamment lexicales) du français dans les différents pays de la Francophonie africaine<sup>1</sup>. L'autre relève des activités de recherches liées à

---

<sup>1</sup> Cf. Queffélec & Niangouna 1990, Frey 1996, Queffélec 1997, Benzakour & Gaadi & Queffélec 2000, Naffati & Queffélec 2004 et Nzesse 2009 (pour ne mentionner qu'une partie du nombre considérable d'ouvrages traitant du sujet en question).

la question des variétés de l'anglais à l'échelle mondiale (« World Englishes »)<sup>2</sup>, parmi lesquelles nous portons une attention particulière aux travaux récents de l'angliciste allemand E. W. Schneider qui s'intéressent à l'évolution de l'anglais dans la perspective de ses différents contextes post-coloniaux.

Le « modèle dynamique » proposé par Schneider (2007 : 29ss)<sup>3</sup> aussi bien que les études effectuées dans la lignée de Lafage et Queffélec démontrent que, sur le plan méthodologique, l'analyse adéquate des constellations contextuelles, dans lesquelles évoluent l'anglais et le français comme langues d'une société post-coloniale, nécessite la prise en considération de données d'orientations fort diverses (extractions de corpus numérisés, résultats obtenus au moyen de tests de perception et/ ou de représentation, analyses et catégorisations de discours identitaires, évaluation de faits politiques, socio-économiques et démographiques dans le cadre de schémas « Status / Corpus », tels ceux présentés par Chaudenson & Rakotomalala 2004). Bien que l'un des objectifs principaux du projet AFRIVAR consiste à développer une plate-forme numérique susceptible de synthétiser ces différents types de données, le présent article se concentrera sur les possibilités d'exploitation de larges corpus électroniques, ce qui est bien en phase avec la visée thématique actuelle des recherches effectuées par l'équipe de Cologne<sup>4</sup>, à savoir la tentative de discerner de façon exhaustive les phénomènes linguistiques dont l'existence plaiderait en faveur de l'hypothèse d'une émergence de nouvelles « variétés nationales » dans les pays francophones d'Afrique et donc du remplacement graduel de la norme acrolectale / hexagonale (exogène) par une ou plusieurs norme(s) mésolectale(s) / locale(s) (endogène[s]) du français en Afrique.

La méthodologie qui sous-tend notre analyse des particularités linguistiques possibles s'inspire en grande partie de la notion de « profil combinatoire » telle qu'elle a été exposée par Blumenthal (2006) et Diwersy (2007).

Certes, l'idée de prendre en considération les faits de combinatoire lexicosyntaxique dans l'étude des variétés d'une langue n'est pas nouvelle. Elle a été mise en avant, en l'occurrence par rapport au français en Afrique, dans un article publié par Stein (2003)<sup>5</sup>, aussi bien que dans plusieurs travaux de Speelman et Geeraerts concernant le néerlandais aux Pays-Bas et en Belgique (cf. Peirsman & Geeraerts & Speelman (2010), l'une des dernières publications en date sur le sujet).

<sup>2</sup> Cf. entre autres Crystal 1997, Kachru & Kachru & Nelson 2006 ainsi que Mesthrie & Bhatt 2008.

<sup>3</sup> Ce modèle dénommé «The Dynamic Model of the evolution of Postcolonial Englishes» distingue cinq phases évolutives, à savoir les phases de fondation (angl. : *foundation*), de stabilisation exonormative (*exonormative stabilization*), de nativisation (*nativization*), de stabilisation endonormative (*endonormative stabilization*) et de différenciation (*differentiation*). Chacune des phases se caractérise, pas forcément de manière symétrique, par quatre paramètres se référant aux contextes des faits historiques et politiques, des constructions identitaires, des aspects sociolinguistiques (contact, usages, fonctions, attitudes) et de l'évolution interne des systèmes phonétique, grammatical et lexical.

<sup>4</sup> Cf. Blumenthal (2011).

<sup>5</sup> Nous partageons notamment les idées avancées par Stein (2003, 14s) en faveur d'une exploitation fréquentielle extensive de corpus compilés dans le but d'analyser les divergences de norme observables dans les différentes régions de la Francophonie.

Pourtant, les différentes approches qu'adoptent les auteurs mentionnés dans l'analyse de la combinatoire n'ont pas forcément les mêmes implications méthodologiques et théoriques que celle que nous défendons : alors que Stein s'inspire en partie d'une modélisation phraséologique avec une attention particulière portée à la notion de « collocation » (2003 : 2s), nous partageons avec Peirsman & Geeraerts & Speelman une orientation clairement distributionnaliste, à cette différence près que nous partons du principe que la combinatoire spécifique d'un item lexical ou grammatical relève de sa valeur systémique et fonctionnelle, et non pas seulement de sa substance conceptuelle<sup>6</sup>.

Ainsi, notre approche s'apparente surtout à la visée méthodologique et théorique du contextualisme britannique tel qu'il est représenté par le courant de la « grammaire lexicale », fondée sur les idées de Sinclair relatives au « principe de l'idiome » (de l'anglais : *idiom principle*) (cf. 1991 ; 2004), ainsi que par la théorie du *lexical priming* (fr. : « amorçage lexical ») développé par Hoey (2005). C'est bien des différents « types d'amorçage » postulés parmi les hypothèses de ce dernier (cf. 2005 : 13) que nous déduisons le modèle descriptif qui sert de base à l'analyse des phénomènes linguistiques dans le cadre du projet AFRIVAR. Si l'on conjugue la notion de « profil combinatoire » et les niveaux cooccurrentiels correspondant aux amorçages collocationnels, sémantiques et colligationnels évoqués par Hoey, l'on obtient le modèle suivant :

**Tableau 1.** Niveaux cooccurrentiels (selon Hoey 2005) et schématicité distributionnelle

Niveau cooccurrentiel	Niveau de schématicité distributionnelle
cooccurrences lexicales	Profil combinatoire
associations sémantiques / pragmatiques	
colligations	
collocations textuelles (chaînes lexicales)	Profil d'intégration textuelle
associations sémantiques textuelles (argumentatives)	
colligations textuelles	

Notre objectif est donc de repérer des spécificités d'usage liées à des patrons cooccurrentiels divergents aux niveaux lexical, syntaxique et textuel (visée descriptive), en vue d'établir, sur la base de ses schémas distributionnels spéci-

<sup>6</sup> Suivant une interprétation structuraliste de Firth (1957 :196), selon qui « [m]eaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words [...] », l'analyse de la combinatoire spécifique d'un item lexical ou grammatical contribue donc en premier lieu à identifier sa signification et non pas les concepts qu'il désigne, comme le suggère l'approche – bien établie dans le domaine de la linguistique informatique – des « Word Space Models » (cf. Sahlgren 2006), dont s'inspirent Peirsman & Geeraerts & Speelman (cf. 2010, 471s).

fiques, la valeur fonctionnelle d'un item donné selon les variétés prises en considération (visée méthodologique et théorique).

La réalisation de cet objectif va de pair avec la mise en œuvre d'une stratégie d'analyse s'inspirant d'une approche à dominante inductive basée sur des ressources numériques et les méthodes de la linguistique de corpus contemporaine, qui, avec des techniques d'exploration telles que la comparaison de données fréquentielles, l'extraction d'affinités cooccurentielles et la classification automatique, peuvent fournir des données susceptibles de compléter les résultats obtenus par la lexicographie différentielle (*cf. supra*) au moyen d'une fouille « classique » d'occurrences. Ces aspects seront concrétisés dans la section suivante qui nous permettra d'aborder le cadre méthodique d'une étude pilote dont nous présenterons les détails dans la quatrième partie du présent article.

### 3. Corpus et outils lexicométriques

Les données lexicométriques, qui font l'objet des paragraphes suivants, sont issues d'une base textuelle construite et mise à jour par l'équipe du CIFRA, au Département des langues romanes de l'Université de Cologne. Le volet journalistique de cette base, le *Corpus numérisé de la presse francophone* (Diwersy 2009-), comprend actuellement environ 1.400.000 articles originaires d'une quinzaine de pays. Le tableau 2 présente les totaux de textes et de mots-occurrences compilés des sous-échantillons correspondants :

**Tableau 2.** Corpus numérisé de la presse francophone – totaux des textes et mots-occurrences compilés par pays

Pays	Nombre de textes compilés	Mots-occurrences
France	619.000	309.500.000
Algérie	91.200	45.600.000
Belgique	49.500	24.750.000
Bénin	19.200	9.600.000
Burkina Faso	16.700	8.350.000
Congo (R.C.)	17.900	8.950.000
Congo (R.D.C.)	38.400	19.200.000
Côte d'Ivoire	42.600	21.300.000
Cameroun	91.200	45.600.000
Canada (Québec)	63.000	31.500.000
Liban	53.200	26.600.000
Madagascar	20.500	10.250.000
Mali	41.300	20.650.000
Maroc	87.400	43.700.000
Sénégal	36.200	18.100.000
Suisse	52.600	26.300.000
Tunisie	93.500	46.750.000
TOTAUX :	1.433.400	716.700.000

À ce jour, plus de la moitié des documents disponibles a été soumise à un traitement automatique comprenant plusieurs étapes, telles que

- le balisage XML d'éléments de structuration textuelle (articles, paragraphes, phrases) et l'identification de méta-données (titre, auteur, date, etc.) ;
- la segmentation de mots-formes ;
- l'annotation morphosyntaxique et la lemmatisation des mots-formes segmentés ;
- le balisage de relations de dépendance syntaxique<sup>7</sup>.

Les textes annotés sont accessibles dans le cadre d'une application basée sur le moteur de requêtes CQP / CWB<sup>8</sup> pour l'extraction de concordances et de données cooccurentielles ainsi que sur le logiciel GNU R<sup>9</sup> pour l'implémentation des calculs statistiques.

Cette application met en place plusieurs procédés lexicométriques, parmi lesquels figurent notamment :

- la détermination d'affinités combinatoires (collocationnelles et colligationnelles) au moyen de certaines mesures d'associations telles que le test du *log-likelihood* (cf. Dunning 1993) et le regroupement des données cooccurentielles sous forme de lexicogrammes (cf. Heiden 2004) ;
- la caractérisation de la répartition interne des lexicogrammes au moyen d'indices de stéréotypicité (cf. Diwersy 2007) ;
- le calcul de scores de similarité entre lexicogrammes (cf. Blumenthal & Diwersy & Mielebacher 2005) ;
- la modélisation des similarités cooccurentielles au moyen de différentes méthodes comme l'échelonnement multidimensionnel ou le partitionnement de données (*clustering*).

Pour notre étude pilote, nous avons créé des échantillons équilibrés par pays en fonction de la quantité de documents annotés disponibles. Le tableau suivant donne un aperçu des sous-corpus respectifs :

---

<sup>7</sup> Les opérations de segmentation, d'annotation morphosyntaxique, de lemmatisation et de balisage des relations de dépendance syntaxique ont été effectuées au moyen du logiciel Connexor (cf. Tapanainen & Järvinen 1997).

<sup>8</sup> Cf. < <http://cwb.sourceforge.net> > (date de consultation : 12/04/2011).

<sup>9</sup> Cf. < <http://www.r-project.org> > (date de consultation : 12/04/2011).

**Tableau 3.** Sous-échantillons pris en considération dans le cadre de l'étude pilote

Code échantillon <sup>10</sup>	Pays	Mots-occurrences
FRA	France	20.578.600
COD	Congo (R.D.C.)	19.245.821
CIV	Côte d'Ivoire	18.873.204
CAM	Cameroun	19.606.634
MAR	Maroc	19.907.421
SEN	Sénégal	20.533.726
TUN	Tunisie	19.313.709

En partant d'une liste de lexèmes établie par l'équipe de recherche EMOLEX<sup>11</sup>, nous avons sélectionné une dizaine de substantifs sémantiquement apparentés, à savoir *peur*,  *Crainte*, *souci*, *appréhension*, *angoisse*, *préoccupation*, *panique* et *inquiétude*. Pour chacun des noms<sup>12</sup> sélectionnés comme pivots, nous avons d'abord créé des lexicogrammes par sous-échantillon en retenant les verbes recteurs cooccurrent avec une fréquence supérieure à 4 et un score d'association supérieur à 10,83 selon *log-likelihood*<sup>13</sup>. Les données cooccurrentielles obtenues grâce à ces opérations lexicométriques sont illustrées par les extraits des lexicogrammes suivants :

<sup>10</sup> La plupart des codes indiquant les différents sous-échantillons ont été conçus dans la lignée des codes de pays prévus par la norme ISO 3166-1 alpha-3.

<sup>11</sup> Cf. < <http://emolex.eu> > (date de consultation: 12/04/2011).

<sup>12</sup> Les unités de requête correspondantes ont été définies par la conjonction d'une indication de lemme et de classe lexicale.

<sup>13</sup> Ceci correspond à une vraisemblance d'erreur de 0,001 selon la répartition du X<sup>2</sup> (cf. Manning & Schütze 2000:609s), à laquelle s'applique le test de significativité *log-likelihood*.

**Tableau 4.** Extraits des lexicogrammes obtenus pour les substantifs *crainte* et *peur* dans les échantillons hexagonal, congolais, camerounais et sénégalais

<b>Pivot</b>	<b>Collocatif</b>	<b>Score <i>log-likelihood</i></b>
crainte_FRA	dissiper_V_H	41,0024
...	...	...
<b>Pivot</b>	<b>Collocatif</b>	<b>Score <i>log-likelihood</i></b>
...	...	...
crainte_RDC	apaiser_V_H	57,944
...	...	...
<b>Pivot</b>	<b>Collocatif</b>	<b>Score <i>log-likelihood</i></b>
...	...	...
peur_CAM	avoir_V_H	865,162
...	...	...
<b>Pivot</b>	<b>Collocatif</b>	<b>Score <i>log-likelihood</i></b>
...	...	...
peur_SEN	caler_V_H	20,4814
...	...	...

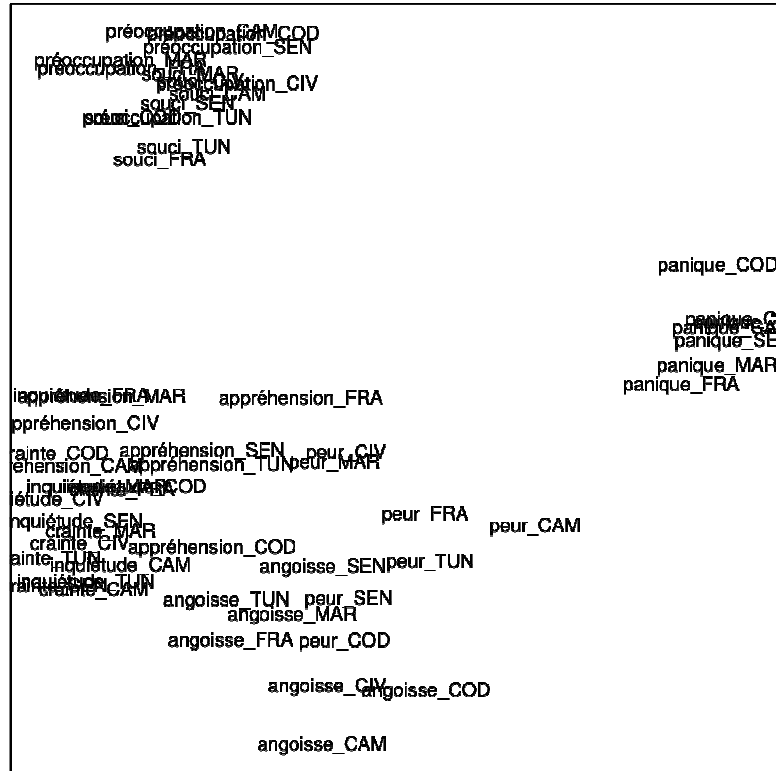
Les lexicogrammes établis pour chaque pivot nominal par rapport aux différents échantillons pris en considération ont ensuite été soumis à des calculs de similarité cosinus basés sur les rangs normalisés des scores *log-likelihood*.<sup>14</sup> À partir du tableau des indices de similarité obtenus, nous avons finalement effectué, par l'intermédiaire d'une transformation en matrice à distances, un échelonnement multidimensionnel ainsi qu'un partitionnement par regroupement hiérarchique pour modéliser les divergences des affinités combinatoires des substantifs sélectionnés.

#### 4. Convergences et divergences des affinités combinatoires

Cette section est consacrée à la présentation et à l'interprétation des tendances observées au moyen des procédés lexicométriques mis en œuvre. Commençons par l'analyse des nuages de points réalisés par échelonnement multidimensionnel des données cooccurentielles prises en considération :

<sup>14</sup> Les calculs en question sont décrits par Blumenthal & Diwersy & Mielebacher 2005.

**Figure 1.** Nuages de points résultant d'un échelonnement multidimensionnel appliqué aux données cooccurentielles retenues. NB : le fait que certains mots du tableau se recouvrent tout en entravant la lisibilité ne porte nullement atteinte à l'objectif de notre démonstration



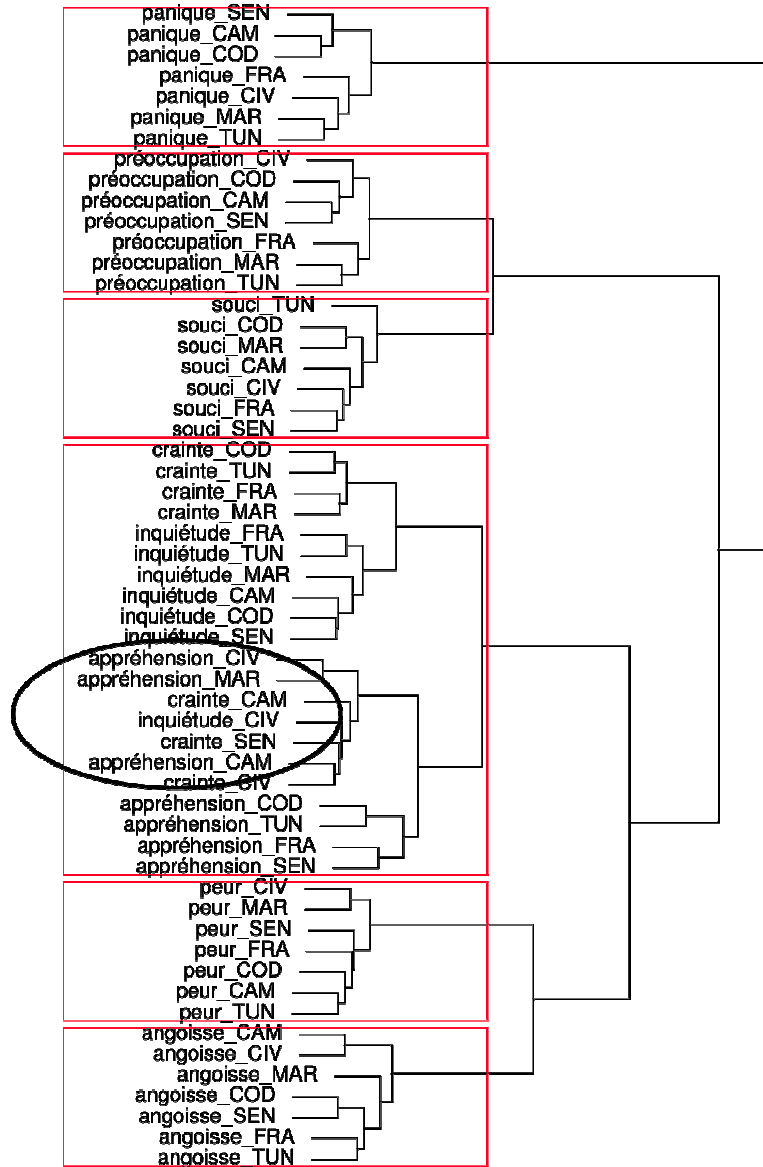
Le graphique en question révèle l'existence de plusieurs ensembles isolés rassemblant *grosso modo* les mêmes items lexicaux indépendamment de leur échantillon d'origine.

Pourtant, il n'y a pas qu'homogénéité. Comme le suggère le dendrogramme résultant d'un partitionnement hiérarchique<sup>15</sup> des données cooccurentielles (*cf.* figure 2), qui nous procure une visualisation plus précise de la composition des différents groupes d'items nominaux, il y a une séparation assez nette concernant les substantifs *appréhension*, *inquiétude* et *crainte* par rapport aux instances respectives provenant des échantillons camerounais (*appréhension*, *crainte*), ivoirien (*appréhension*, *crainte*, *inquiétude*), marocain (*appréhension*) et sénégalais (*crainte*).

<sup>15</sup> L'opération de partitionnement et la visualisation des résultats sous forme de dendrogramme ont été effectuées au moyen du module R *hclust* (*cf.* < <http://stat.ethz.ch/R-manual/R-devel/library/stats/html/hclust.html> >, date de consultation : 14/04/2011) sur la base de l'algorithme *ward*.



Figure 2. Partitionnement agglomératif hiérarchique des données cooccurentielles retenues



Cette observation se confirme, lorsqu'on prend en considération la répartition des scores de similarité dans l'intervalle regroupant les premiers rangs des populations établies pour chacun des substantifs en question.

Alors qu'on obtient par rapport à *panique* un taux de convergence maximal dans l'intervalle des 21 premiers rangs possibles sur l'ensemble des couples formés

par les différentes instances d'un même substantif et que ce taux reste fort élevé dans le cas des noms *préoccupation*, *souci* et *peur*, on retrouve à l'autre bout de l'échelle *appréhension* avec  *Crainte*, *inquiétude* et *angoisse* en position moyenne :

**Tableau 5.** Taux de convergence dans l'intervalle des premiers rangs

Nom	Nombre de couples dans l'intervalle	Nombre de positions dans l'intervalle	Taux de convergence (%)
panique	21	21	100,00
préoccupation	20	21	95,24
souci	19	21	90,48
peur	18	21	85,71
angoisse	13	21	61,9
Crainte	11	21	52,38
inquiétude	11	21	52,38
appréhension	4	21	19,05

Vu l'écart considérable signalé par le tableau précédent, il semble pertinent de comparer les profils combinatoires obtenus pour *appréhension* selon les différents échantillons pris en compte. Dans l'analyse comparative des préférences cooccurrentielles respectives, nous allons nous concentrer sur les associations sémantiques telles qu'elles sont réalisées par les collocatifs verbaux spécifiques du substantif-pivot en question.

Avant d'entrer dans les détails, il est indispensable de présenter, ne serait-ce que de manière sommaire, les éléments classificatoires qui entrent en jeu dans la comparaison des dimensions sémantiques faisant l'objet d'une variation dans le contexte des profils cooccurrentiels d'*appréhension* différenciés par échantillon. En nous inspirant en partie de la grille d'analyse élaborée dans le cadre du projet EMOLEX, nous distinguerons trois classes majeures de verbes collocatifs, à savoir :

1. les **verbes à orientation situationnelle** se référant au mode d'existence de l'état des choses désigné par le nom d'affect, qu'il s'agisse d'une mise en perspective absolue (ayant trait à l'expérientiation en tant que telle) ou scalaire (ayant trait à l'intensité). En ajoutant les variantes à opérateur causal ou de cyclicité (indiquant une reprise), on retiendra les sous-classes suivantes :

- verbes supports neutres (Sit:Exist:pos : *avoir*, *ressentir*)
- verbes supports phasiques : inchoatifs (Sit:Exist:inch : *naître*), continuatifs (Sit:Exist:cont : *perdurer*) ou terminatifs (Sit:Exist:term : *disparaître*)
- verbes de causation positive (Caus:Sit:Exist:pos : *susciter*) ou négative (Caus:Sit:Exist:neg : *dissiper*)
- verbes de causation de reprise (Caus:Sit:Exist:pos:rep : *réveiller*)

2. les **verbes à orientation évidentielle** se référant au repérage de l'affect constaté. À cet égard, on distinguera les sous-classes suivantes :

- verbes de manifestation (Evid:manif : *exprimer*, *émettre*, *afficher*, *manifester*)

- verbes de repérage par indices (Evid:indic : *témoigner, trahir*)

3. les **verbes à orientation évaluative** se référant à un jugement de pertinence concernant l'attitude épistémique attribuée à l'expérient. Dans notre analyse, nous tiendrons compte des sous-classes suivantes :

- verbes d'évaluation convergente rapportée (Eval:pert:pos:indir : *comprendre*) ou assumée par le locuteur (Eval:pert:pos:dir : *se confirmer, justifier*)
- verbes d'évaluation divergente rapportée (Eval:pert:neg:indir : *étonner*)<sup>16</sup>

Notons d'emblée que le seul collocatif verbal spécifique du nom *appréhension* à travers l'ensemble des échantillons considérés est le verbe support neutre *avoir* :

- (1) Il est convaincu que le prix FDI va énormément contribuer à donner de la confiance aux investisseurs étrangers qui **ont** encore des **appréhensions**. (CIV, *Fraternité Matin*, 2007)
- (2) Elles [les entreprises françaises] ne doivent donc pas **avoir** d'**appréhensions** par rapport à la concurrence des pays arabes, de l'Inde, de la Chine quand on sait qu'une entreprise française a perdu un marché du port de Dakar au profit de celle d'un pays arabe ? (SEN, *Walf Fadjri*, 2008)

Si ce verbe sert à établir une mise en perspective situationnelle dans le cadre de tous les profils combinatoires relevés pour *appréhension*, c'est bien par rapport à cette dimension sémantique qu'on observe également quelques divergences qui sont à l'origine de l'opposition d'échantillons constatée plus haut. Les divergences distributionnelles en question concernent notamment les verbes de causation positive et de causation négative, dont les affinités combinatoires se focalisent sur un nombre restreint d'échantillons selon deux configurations respectives.

Dans le cas des verbes de causation positive, on peut constater que leur statut de collocatifs spécifiques du nom *appréhension* se limite, du moins dans le cadre de notre corpus de travail, aux échantillons camerounais, marocain et congolais :

- (3) Pesage routier : Des transporteurs de l'Adamaoua inquiets – Le démarrage officiel des activités du poste de pesage de Mawui vendredi dernier **suscite** des **appréhensions**. (CAM, *Mutations*, 2007)
- (4) Le spectre de la crise des « subprimes » aux USA a pu, à un moment donné, **susciter** certaines **appréhensions** chez les divers acteurs voire chez les acquéreurs. (MAR, *Le Matin du Sahara*, 2008)
- (5) Ces propos, [...], n'en expriment pas moins des craintes à caractère politique toujours actuelles dont nous devons tenter d'entrevoir les causes en République démocratique du Congo, en examinant à présent l'**appréhension** politique que **suscite** l'intérim du Président de la République, [...], par le président du Sénat. (COD, *Le Potentiel*, 2007)

<sup>16</sup> Dans le cadre du corpus de travail que nous avons consulté, nous n'avons pas pu relever, au-dessus du seuil de spécificité statistique défini, des verbes exprimant une évaluation divergente assumée par le locuteur, ce concept étant plutôt réalisé au moyen de la négation *ne... pas* ajoutée à un verbe d'évaluation convergente tel *se confirmer*.

Pour ce qui est de l'affinité combinatoire d'*appréhension* par rapport à la classe des verbes de causation négative, elle est à la base d'une configuration qui regroupe aussi bien l'échantillon camerounais que les échantillons ivoirien et tunisien :

- (6) Pour **dissiper** les **appréhensions** de certains conducteurs qui craignent de se voir retirer le permis de conduire au cas où ils seraient jugés inaptes du fait du modulo test, les concepteurs du logiciel répondent qu'il est possible de réhabiliter les chauffeurs défaillants, grâce à d'autres modules [...]. (CAM, *Mutations*, 2007)
- (7) Ce fut l'occasion de **dissiper** les **appréhensions** des militants et de lever les inquiétudes liées à la perte de pièces d'identité pendant la guerre. (CIV, *Fraternité Matin*, 2007)
- (8) Les réponses obtenues étaient floues et n'avaient pas **dissipé** les **appréhensions** du jeune homme. (TUN, *Le Temps*, 2008)

On note par ailleurs que le profil combinatoire d'*appréhension* extrait du corpus camerounais intègre à lui seul l'opposition entre verbes de causation positive et de causation négative ; dans le cadre de l'échantillon en question, cela renforce sa proximité paradigmatique avec *inquiétude* aux dépens de *crainte*<sup>17</sup> – et ceci d'autant plus que, dans le cas de l'échantillon camerounais, à la différence de *crainte*, les combinatoires spécifiques d'*appréhension* et d'*inquiétude* convergent, toutes deux incluant la classe des verbes continuatifs<sup>18</sup>, qui, comme *subsister* par exemple, entretiennent un lien inférentiel étroit avec les verbes de causation négative<sup>19</sup> :

- (9) Mais au-delà de l'allégresse, quelques **appréhensions subsistent**. Les populations s'inquiètent sur leur sécurité sur ce pont. (CAM, *Cameroon Tribune*, 2007)
- (10) En attendant la fin des travaux de déviation qui s'annonce imminente, ils n'éprouvent plus assez de difficultés pour traverser la zone du glissement. Mais des **inquiétudes persistent**. (CAM, *La Nouvelle Expression*, 2008)

Mis à part les verbes à orientation situationnelle, nous allons considérer par la suite les deux classes de collocatifs verbaux jouant un rôle crucial dans la constitution des groupements d'échantillons qu'on peut observer selon la variation des affinités combinatoires d'*appréhension*. À chaque fois, on placera au centre de nos considérations le groupe « dissident » des profils combinatoires d'*appréhension* établis sur la base des échantillons camerounais, ivoirien et marocain, dont le statut particulier est indiqué par la figure 2.

D'une part, l'un des éléments fondateurs de la configuration qui rassemble les variantes camerounaise, ivoirienne et marocaine du profil cooccurentiel d'*ap-*

<sup>17</sup> Hormis le cas de l'échantillon camerounais, c'est bien le profil combinatoire de *crainte*, et non pas celui d'*appréhension*, qui présente les taux de similarité les plus élevés par rapport à *inquiétude*.

<sup>18</sup> L'affinité combinatoire d'*inquiétude* à la classe des verbes supports continuatifs est bien attestée dans l'ensemble des échantillons considérés.

<sup>19</sup> On constate la présence de ce lien inférentiel en ce sens que les verbes continuatifs présupposent l'impact d'une force susceptible de mettre fin à la situation dénotée par le nom qu'ils supportent. Alors que ceux-ci permettent d'affirmer l'échec de cet impact, les verbes de causation négative transportent l'assertion de l'état de non-existence qui lui succède.

*préhension*, relève de l'affinité combinatoire prononcée pour les verbes évidentiels, et notamment pour la série de collocatifs se référant, comme leur représentant prototypique *exprimer*, à la verbalisation de l'affect désigné par le substantif-pivot. S'ajoute à cette configuration illustrée ci-dessous le profil combinatoire d'*appréhension* issu de l'échantillon tunisien bien que, pour ce dernier, le poids revenant à la classe de verbes en question reste fort réduit :

- (11) À ce sujet, un responsable en service à la délégation provinciale du Minader du Sud n'a pas manqué d'**exprimer** quelques **appréhensions**. (CAM, *La Nouvelle Expression*, 2007)
- (12) Le Forum national sur la dette et la pauvreté (FNDP) redoute les accords de partenariat économique (APE). Il a **exprimé** ses **appréhensions**, hier, à son siège, à Abidjan-Adjamé, lors d'une conférence de presse, animée par son secrétaire exécutif, Diakalia Ouattara. (CIV, *Notre Voie*, 2008)
- (13) [...] soulignant que le Maroc dispose d'un statut mondialement reconnu dans cette discipline [le taekwondo]. M. Ismaïli a toutefois **exprimé** son **appréhension** que ses protégés tombent sur les grandes nations dès les premiers tours [...]. (MAR, *Aujourd'hui Le Maroc*, 2008)

Si la configuration regroupant les variantes camerounaise, ivoirienne et marocaine du profil combinatoire d'*appréhension* se base largement sur la présence de collocatifs spécifiques appartenant à la classe des verbes de manifestation, elle se délimite, d'autre part, par l'absence simultanée de la classe des verbes d'évaluation affirmative, qui font bien partie intégrante de la combinatoire préférentielle d'*appréhension*, telle qu'elle se présente dans le groupe complémentaire d'échantillons (à l'exception du corpus hexagonal). Notons à l'égard des affinités cooccurrentielles d'*appréhension* impliquant d'autres classes de verbes liées à la dimension de l'évaluation propositionnelle que le profil combinatoire issu de l'échantillon camerounais est le seul à présenter, avec *étonner*, un collocatif se référant à une évaluation négative :

- (14) Ceci est un dossier sensible, j'en conviens, et je crois qu'il ne vaudrait pas mieux mélanger ici les données de l'histoire avec des **appréhensions** non **justifiées**. (COD, *Le Potentiel*, 2008)
- (15) [...] les clubs sénégalais engagés sur le front africain [...] allaient au-devant de grosses difficultés. La première manche des seizièmes de finale de la ligue des champions et de la coupe de la Confédération a largement **confirmé** cette **appréhension**. (SEN, *Le Soleil*, 2008)
- (16) À notre sens, rien ne **justifiait** l'**appréhension** et la prudence excessives de l'entraîneur aghlabide, frileux et un peu trop timoré. (TUN, *La Presse*, 2008)
- (17) Avec un optimisme débordant, Alioum croit en la victoire des Lions et **s'étonne** de l'**appréhension** affichée par quelques Camerounais de Maroua. (CAM, *Mutations*, 2008)

En guise de résumé, le tableau ci-dessous donne un aperçu des classes de verbes qui déterminent, en termes de dimensions sémantiques, la variation de la combinatoire spécifique d'*appréhension* selon les échantillons pris en compte. Outre le caractère fort diversifié des relations de convergence et de divergence variationnelles qu'entretient le profil combinatoire d'*appréhension* provenant de

l'échantillon camerounais, ce tableau montre également la sous-détermination relative des affinités cooccurentielles du même nom dans le cas du corpus hexagonal. D'un point de vue quantitatif, cette sous-détermination s'explique par le faible taux de fréquence que présente *appréhension* dans le cadre de l'échantillon hexagonal<sup>20</sup> et la diminution du nombre potentiel de collocatifs spécifiques qui en résulte.

**Tableau 6.** Distribution des dimensions sémantiques du profil combinatoire d'*appréhension* selon les échantillons considérés

Dimensions sémantiques et classes de verbes <sup>21</sup>		Échantillons source du profil combinatoire d' <i>appréhension</i>						
		CAM <sup>22</sup>	CIV	COD	FRA	MAR	SEN	TUN
SIT	Exist	avoir	avoir	avoir	avoir	avoir	avoir	avoir
	Caus:pos	susciter	-	susciter	-	susciter	-	-
	Caus:neg	dissiper	dissiper	-	-	-	-	dissiper
	Exist:cont	subsister	-	-	-	-	-	-
EVID	Evid:manif	exprimer	exprimer	-	-	exprimer	-	exprimer
EVAL	Eval:pert:pos:dir	-	-	justifier	-	-	confirmer	justifier
	Eval:pert:neg:indir	s'étonner	-	-	-	-	-	-

Le statut particulier de l'échantillon camerounais que révèle l'aperçu de la combinatoire spécifique d'*appréhension* permettrait certes une exploration approfondie et il serait sans doute tentant d'établir une corrélation avec le fait qu'au Cameroun, le français coexiste dans le domaine fonctionnel de la communication écrite

<sup>20</sup> À cet égard, on peut constater que le taux de fréquence relative ( $3 \cdot 10^{-6}$  pour une fréquence absolue de 62 occurrences) qui revient à *appréhension* dans le cadre du corpus hexagonal est nettement inférieur aux chiffres respectifs obtenus sur la base du reste des échantillons pris en considération. Selon le test des « mots-clés » (cf. Rayson 2003 ; Scott & Tribble 2006), qui donne un score *log-likelihood* de -52,95, ce taux se caractérise par une sous-représentativité hautement significative compte tenu de la distribution fréquentielle globale. Il reste à noter que la fréquence relative d'*appréhension* dans le corpus camerounais se retrouve à l'autre bout de l'échelle avec un taux largement excédentaire de  $1,1 \cdot 10^{-5}$  (pour une fréquence absolue de 212), ce qui correspond à une sur-représentativité significative bien attestée par une valeur *log-likelihood* s'élevant à 55,61.

<sup>21</sup> Les codes indiquant les classes de verbes ont été introduits par la présentation se situant à la page 11 de cet article.

<sup>22</sup> Pour les codes échantillon employés, voir la liste donnée par le tableau 3.

avec l'anglais, c'est-à-dire une langue qui présente avec *apprehension* un lexème étymologiquement apparenté, lequel se prêterait peut-être, de par ses affinités combinatoires, à des transferts constructionnels. Mais une telle étude se heurterait, dès le départ, à deux problèmes d'ordre méthodique. D'une part, nous ne disposons pas pour l'instant, malgré l'existence du *Corpus of Cameroon English*<sup>23</sup>, d'un échantillon comparable en termes de quantité aussi bien que de possibilités d'exploitation lexicométrique selon les procédés présentés dans cet article. D'autre part, on notera l'absence, dans l'immédiat, d'un dispositif susceptible de corrélérer les tendances relevées au niveau des spécificités combinatoires et les contextes acquisitionnel et sociétal des locuteurs camerounais. À cet égard, on dispose, grâce à Mendo Zé (2009), de données adéquates auxquelles il faudrait tout de même conjuguer les données de corpus dans le cadre d'une expérimentation qui, au bout du compte, reste à développer.

## 5. Conclusion

Dans le présent article, nous avons montré les possibilités qu'offre une approche à dominante inductive – c'est-à-dire, guidée par l'observation de patrons cooccurrentiels – pour l'analyse comparative de la variation du français écrit en France et dans différents pays de la francophonie africaine. Dans ce but, nous avons élaboré un cadre méthodique comprenant une série de procédés lexicométriques appliqués à un vaste corpus de presse francophone divisé en échantillons délimités selon le critère de provenance géographique. L'étude pilote que nous avons menée sur la combinatoire lexico-syntaxique spécifique d'un groupe de lexèmes sémantiquement proches (à savoir des noms d'affect appartenant aux champs notionnels 'peur' / 'souci' / 'inquiétude') nous a ensuite permis de détecter quelques écarts variationnels importants parmi lesquels nous avons relevé le cas exemplaire du profil combinatoire d'*apprehension* qui présentait un certain nombre de divergences distributionnelles selon les dimensions sémantiques établies par ses affinités cooccurrentielles respectives. On a pu relever à cet égard l'extension fonctionnelle accrue d'*apprehension* dans le cas de l'échantillon camerounais ; le profil de ce mot est formé d'un ensemble d'associations sémantiques qui s'avèrent plus dispersées parmi le reste des échantillons considérés. Ainsi, en ce qui concerne la dimension de mise en perspective situationnelle, la combinatoire spécifique d'*apprehension*, telle qu'elle émane de l'échantillon camerounais, est la seule à intégrer l'opposition entre verbes de causation positive et de causation négative et à présenter des verbes supports continuatifs. Ceci va de pair avec une proximité paradigmatique élevée envers *inquiétude* au détriment de *crainte*.

L'approche lexicométrique exposée, dont nous croyons avoir démontré l'utilité, se prête elle-même à la conception d'un programme méthodique élargi qui intégrerait – en aval de la description des données induites sur la base de vastes cor-

<sup>23</sup> Le projet de corpus en question, qui évolue depuis plusieurs années sous la direction de D. Nkemleke, est hébergé par les Départements d'anglais de l'Université de Chemnitz ainsi que de l'École Normale Supérieure à l'Université Yaoundé I (cf. < <http://www.tu-chemnitz.de/phil/english/ccf/> > ; < <http://www.ens.cm/spip.php?article295> >, date de consultation : 12/09/2011).

pus – une phase d’exploitation expérimentale effectuée au moyen de tests de perception à orientation psycholinguistique (visant le traitement perceptif « tacite » d’une séquence lexico-syntaxique spécifique) et sociolinguistique (visant les représentations perceptives explicites). Un tel programme, dont le but serait de mettre en évidence, dans une perspective complémentaire, le rapport entre les divergences structurales observées, la sensibilité idiomatique des locuteurs et leurs représentations de la norme, pourrait représenter un élément constitutif dans l’exploration des faits variationnels en francophonie, en termes du modèle de l’évolution cyclique formulé par Schneider (2007). Histoire à suivre...

### Bibliographie

- BENZAKOUR, F. & DRISS, G. & QUEFFÉLEC, A. (2000). *Le français au Maroc. Lexique et contacts de langues*. Bruxelles, De Boeck & Larcier.
- BLUMENTHAL, P. (2006). *Wortprofil im Französischen*. Tübingen, Niemeyer.
- BLUMENTHAL, P. (2011). « Les verbes dans la presse francophone d’Afrique noire », in *Le français en Afrique* n°26, pp. 117-136.
- BLUMENTHAL, P. & DIWERSY, S. & MIELEBACHER, J. (2005). « Kombinatorische Wortprofile und Profilkontraste. Berechnungsverfahren und Anwendungen », in *Zeitschrift für romanische Philologie* n°121, pp. 49-83.  
< [http://www.romanistik.uni-koeln.de/home/blumenthal/publications/wortprofil\\_zrph121.pdf](http://www.romanistik.uni-koeln.de/home/blumenthal/publications/wortprofil_zrph121.pdf) >
- CHAUDENSON, R. & RAKOTOMALALA, D. (2004). *Situations linguistiques de la Francophonie : État des lieux*. Agence Universitaire de la Francophonie (AUF).
- CRYSTAL, D. (1997). *English as a Global Language*. Cambridge, CUP.
- DIWERSY, S. (2007). *Kookkurrenz, Kontrast, Profil. Studien zur lexikalisch-syntaktischen Kombinatorik französischer Substantive*. Thèse de doctorat. Universität zu Köln.
- DIWERSY, S. (2009-). *Corpus numérisé de la presse francophone. Kölner Romanistische Korpusdatenbank*, Köln, Romanisches Seminar der Universität zu Köln.
- FIRTH, J. R. (1957). « Modes of Meaning », in Firth, J. R. (éd.), *Papers in Linguistics 1934-1951*, London e. a., OUP, pp. 190-215.
- FREY, C. (1996). *Le français au Burundi : lexicographie et culture*. Vanves, EDICEF.
- HEIDEN, S. (2004). « Interface hypertextuelle à un espace de cooccurrences: implémentation dans Weblex », in Purnelle, G. & Fairon, C. & Dister, A. (éds.) *Le poids des mots : Actes des 7<sup>e</sup> Journées internationales d’Analyse statistique des Données Textuelles (JADT 04, Louvain-la-Neuve, 10-12 mars 2004)*, Louvain-la-Neuve, Presses Universitaires de Louvain, 2004, pp. 577-588.
- HOEY, M. (2005). *Lexical Priming. A New Theory of Words and Language*, London & New York, Routledge.



- KACHRU, B. B. & KACHRU, Y. & NELSON, C. L. (2006). *The Handbook of World Englishes*. Malden MA. et alii, Blackwell.
- MANNING, C. D. & SCHÜTZE, H. (2000). *Foundations of Statistical Natural Language Processing*. Cambridge MA, MIT.
- MENDO ZÉ, G. (2009). *Insécurité linguistique et appropriation du français en contexte plurilingue*, Paris et alii, L'Harmattan.
- MESTHRIE, R. & BHATT, R. M. (2008). *World Englishes : The Study of New Linguistic Varieties*. Cambridge et alii, CUP.
- NAFFATI, H. & QUEFFÉLEC, A. (2004). *Le français en Tunisie*, Nice, Institut de linguistique française (*Le français en Afrique* n°18).
- NZESSE, L. (2009). *Le français au Cameroun*, Nice, Institut de linguistique française (*Le français en Afrique* n° 24).
- PEIRSMAN, Y. & GEERAERTS, D. & SPEELMAN, D. (2010). « The Automatic Identification of Lexical Variation between Language Varieties », in *Journal of Natural Language Engineering* n° 16 (4), pp. 469-491.
- QUEFFÉLEC, A. & NIANGOUNA, A. (1990). *Le français au Congo (R. P. C.)*. Aix-en-Provence, Université de Provence.
- RAYSON, P. (2003). *Matrix : A Statistical Method and Software Tool for Linguistic Analysis through Corpus Comparison*. Ph. D. thesis, Lancaster University.  
< <http://ucrel.lancs.ac.uk/people/paul/publications/phd2003.pdf> >
- SAHLGREN, M. (2006). *The Word-Space Model : Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*, Ph. D. thesis, Department of Linguistics, Stockholm University.  
< <http://www.sics.se/~mange/TheWordSpaceModel.pdf> >
- SCHNEIDER, E. W. (2007). *Postcolonial English : Varieties around the world*. Cambridge et alii, CUP.
- SCOTT, M. & TRIBBLE, C. (2006). *Textual Patterns : Key Words and Corpus Analysis in Language Education* (Studies in corpus linguistics, 22). Amsterdam et alii, Benjamins.
- SINCLAIR, J. McH. (1991). *Corpus, Concordance, Collocation*. Oxford, OUP.
- SINCLAIR, J. McH. (2004). *Trust the Text : Language, Corpus and Discourse*. London, Routledge.
- STEIN, A. (2003). « Lexikalische Kookkurrenz im afrikanischen Französisch », in *Zeitschrift für französische Sprache und Literatur* n°113, pp. 1-17.
- TAPANAINEN, P. & JÄRVINEN, T. (1997). « A non-projective dependency parser », in *Proceedings of the 5th Conference on Applied Natural Language Processing*. Washington, DC, pp. 64-74.