

Normes occidentales vs usages endogènes camerounais : diagnostic des écarts lexicaux en contexte francophone

Noelle Laura Nguenama^{1 2}, Honoré Blaise Edzougou Afanda⁴, Sandra Invernizzi³, Laurent Lefebvre³

¹ Institut des Sciences Logopédiques, Université de Neuchâtel, Suisse

² Département de Langues Africaines et Linguistique, Université de Yaoundé 1, Cameroun

³ Service de Psychologie Cognitive et de Neuropsychologie, Université de Mons, Belgique

⁴ Département d'Éducation Spécialisée, Université de Yaoundé 1, Cameroun

Abstract. Recourir aux bases de données occidentales pour construire des tests d'aphasie spécifiques aux populations d'Afrique subsaharienne est problématique car, ces normes ne reflètent pas les usages endogènes qui leur sont propres. Cette étude vise à mettre en évidence les écarts lexicaux entre le français camerounais et le français standard. Vingt items lexicaux sur une entrée visuelle – préalablement adaptés au contexte camerounais – ont été exploités et les moyennes issues d'une étude psycholinguistique liée à leur familiarité en contexte ont été comparées à leurs fréquences extraites de Lexique 3. Le logiciel RV4.3 via les coefficients de variation-CV (ET/moyenne x 100) MAE, RMSE, la corrélation (Pearman r), les tests comparatifs (t de Student apparié, t indépendant), la puissance statistique post-hoc (G*Power) et la régression linéaire simple pour R^2 ajusté a permis d'analyser ces écarts. Globalement, il existe une faible corrélation entre ces deux variables ($r \approx 0,25$), des divergences marquées pour les items de moyenne fréquence et des écarts massifs chez les items de haute fréquence. Cette étude révèle des écarts lexicaux considérables entre les cadres camerounais et européen francophones, plaidant ainsi pour la création des bases de données contextualisées en Afrique subsaharienne francophone, afin de développer des tests adaptés à leurs populations.

Mots clés : Français, Cameroun, Normes occidentales, Lexique, Aphasie

1 Introduction

Des estimations récentes de l'Organisation Internationale de la Francophonie (OIF) place la langue française en 5^e position au monde par son nombre de locuteurs – environ 321 millions [1]. Ce rapport de l'OIF stipule que cette langue est présente sur les cinq continents et bénéficie d'une croissance rapide dans les pays d'Afrique subsaharienne - où elle enregistre le plus grand nombre d'utilisateurs quotidiens - notamment au Cameroun. En effet, le Français et l'Anglais sont les deux langues officielles de ce pays, héritage de la double colonisation française et britannique qui marque son histoire. Sa configuration sociolinguistique parmi les plus complexes d'Afrique subsaharienne se caractérise par une richesse linguistique remarquable et des pratiques langagières en constante évolution [2]. Entre 273 et 280 langues nationales coexistent aux côtés du Français et de l'Anglais [3,4]. S'y ajoutent des pratiques langagières hybrides mêlant code-switching et emprunts, à l'exemple du Camfranglais [5]. Ainsi, bien que la majorité des locuteurs camerounais utilisent le français – huit régions sur 10 étant francophones -, ce plurilinguisme vécu dans lequel chaque camerounais parle au moins deux langues [6], influence le Français standard sur le plan culturel [7]. Par ailleurs, dans le cadre d'une étude portant sur la variation linguistique du Français au Cameroun, un auteur [8] propose une analyse pionnière des variétés régionales du Français en contexte camerounais, en mettant en lumière non seulement des catégories sociolinguistiques (telles que des dialectes régionaux, le français commun et les argots) mais aussi une géographie linguistique qui se déploie à travers des grandes zones francophones du pays. En effet, le Cameroun est généralement décrit comme structuré en grandes aires culturelles, notamment la zone côtière (Sawa), les Grassfields (incluant les Bamiléké), l'aire Fang-Béti (Centre, Sud, Est) et la zone soudano-sahélienne au Nord) [9]. Cette stratification linguistique régionale du français au Cameroun est donc bien plus qu'une simple variation de surface car, elle reflète des rapports historiques et identitaires qui complexifient l'usage du Français au Cameroun. Dans ce cadre francophone où l'on retrouve des répertoires plurilingues structurés et socialement hiérarchisés [10, 11], les locuteurs présentent également des profils hétérogènes en termes de niveau d'éducation, d'exposition au Français standard scolaire, et de contact avec les variétés médiatiques ou endogènes du Français. Cette situation influence aussi directement leurs représentations lexicales et leur jugement de familiarité des mots. Ainsi, la familiarité lexicale observée dans les populations camerounaises ne correspond pas nécessairement aux indices de fréquence issus de corpus occidentaux normés, illustrant un décalage entre norme exogène et usages endogènes. Cette perspective rejoint les travaux sur les normes endogènes du Français en Afrique, qui soulignent que les variétés africaines du Français constituent des systèmes linguistiques fonctionnels et contextualisés plutôt que de simples écarts à la norme standard [12].

Cet état des choses rend problématique l'utilisation directe des tests logopédiques et neuropsychologiques francophones importés d'Europe au Cameroun, tests qui sont construits sur la base de la fréquence lexicale du français standard (Ex DTLA & BECLA [13, 14]). Une telle pratique si mise en œuvre risquerait d'entraîner une évaluation langagière biaisée en contexte camerounais. Concrètement, cette évaluation ne prendrait pas en compte les spécificités lexicales et sociolinguistiques du français camerounais, exposant ainsi l'évaluation à des erreurs d'interprétation et à des diagnostics erronés. Cela est davantage problématique puisqu'il existe une carence en évaluation experte dans les domaines de la logopédie et de la neuropsychologie au Cameroun [15]. La nécessité donc de prendre en compte ces particularités du français parlé au Cameroun est pertinente pour permettre l'évaluation fiable et valide des performances langagières dans les aphasies dans ce contexte.

L'aphasie est un trouble acquis du langage résultant d'une lésion cérébrale consécutive à un accident vasculaire cérébral, une maladie neurogénérative, un traumatisme crânien ou une épilepsie [7]. Cette pathologie affecte différentes composantes du langage, notamment les versants de production et compréhension du langage [16, 17]. Des études épidémiologiques internationales indiquent que l'aphasie survient environ chez 25 % à 40 % des patients victimes d'un premier AVC [18]. Dans un pays comme le Cameroun, cela représente potentiellement plusieurs dizaines de milliers de patients post-AVC susceptibles de présenter des troubles du langage au même titre que des patients épileptiques dont les taux dans ce pays sont parmi les plus élevés au monde [19]. Ces troubles langagiers nécessitent donc une évaluation clinique spécialisée. Or, l'absence d'outils de tests d'aphasie adaptés – et validés – aux réalités sociolinguistiques locales limite actuellement la précision diagnostique et la mise en place des thérapies appropriées en Afrique subsaharienne en général et au Cameroun en particulier.

Dans le domaine de la psychologie cognitive et du traitement du langage, des auteurs [20] évoquent des recherches fondamentales concernant la fréquence lexicale (objective) et la familiarité (subjective) des mots dans la construction des tests d'aphasie, notamment lorsque les bases de données lexicales standardisées sont inexistantes. Ils soulignent que, bien que ces deux notions soient souvent corrélées – les mots à forte fréquence étant généralement plus familiers – la familiarité subjective constitue une mesure distincte qui exerce une influence significative sur le traitement lexical et qui vient compléter la mesure de la fréquence objective, entendue comme le nombre d'occurrences d'un mot dans un corpus à l'exemple de lexique 3 [21]. Faute d'une base de données lexicale spécifique – en contexte camerounais – permettant de déterminer les variables psycholinguistiques associées à la fréquence lexicale des items, la présente recherche a utilisé la familiarité lexicale (mesure subjective) comme substitut pour la confronter à leur fréquence lexicale dans des bases de données standards du français européen. Cette procédure est une solution pragmatique dans les contextes d'évaluation linguistique [20]. Ainsi, dans le cadre d'une étude sur le développement d'une batterie d'évaluation de l'aphasie épileptique spécifiquement pensée pour la francophonie camerounaise adulte – dont l'un des sous-test est lié à l'accès lexical en production orale – [15], une étude préalable a permis d'opérer des adaptations linguistico-culturelles de certains items lexicaux en contexte camerounais, notamment pour l'épreuve de dénomination [7]. Sans prétendre à une exhaustivité normative, la présente étude diagnostique a pour objectif de mettre en lumière – à petite échelle – les écarts lexicaux en confrontant la fréquence lexicale des items extraite des bases de données standardisées de la francophonie européenne à la familiarité lexicale desdits items en contexte camerounais.

2 Méthode

Les items lexicaux retenus pour la présente étude sont extraits du sous-test de la Batterie d'Évaluation du Langage de l'Adulte Aphasique Épileptique-BELAE [15]. Cet outil à 100 points constitué de six sous-tests est la toute première batterie conçue pour évaluer spécifiquement l'ensemble des domaines langagiers – en production et compréhension orale – associés aux épilepsies auprès d'une population adulte francophone africaine, et plus particulièrement au sein de la francophonie camerounaise.

2.1 Echantillon et collecte des données de familiarité lexicale

Vingt items (10 noms d'entités concrètes, 10 verbes d'action) adaptés au contexte camerounais [7] ont été retenus. Ils couvraient les domaines liés à l'agriculture, l'alimentation et les actions verbales basiques : *Pomme, lit, chien, ciseaux, fleur, maison, chèvre, balai, banane, scie, attraper, boire, pleuvoir, couper, cultiver, tomber, dormir, nager et jouer*. Ces adaptations ont fait l'objet d'une investigation psycholinguistique de leur familiarité (fa) [15] auprès d'un échantillon unique (30 participants) de genre (15 femmes & 15 hommes), d'âge (12-34 ans ; M : 23,73 ; Écart-type : 3,43) et d'éducation (12-19 années ; M : 15,66, Écart-type : 1,87) varié. Les mesures, recueillies à l'aide d'une échelle en 5 points [22] évaluaient le degré de reconnaissance subjective d'un mot : entre 1 et 2,25 = pas du tout familier ; entre 2,25 et 3,5 = familier ; entre 3,5 et 5 = très familier. Cette collecte a reçu l'aval du Comité National d'Éthique pour la Recherche en Sciences Humaines (CNERSH) du Cameroun.

2.2 Extraction des fréquences lexicales objectives

Les fréquences lexicales (fre, en occurrences par million de mots, pmo) ont été extraites de Lexique 3 [21]. Cette ressource compile environ 140 000 lemmes français issus de deux corpus : a) livres (fre_livres, ~ 31 millions de mots) ; et b) sous titres de films (fre_films ~ 50 millions de mots). Ce deuxième corpus est utilisé ici pour l'extraction des fréquences lexicales car est classiquement utilisé pour les analyses psycholinguistiques. Les seuils - standards [22]. - suivants permettent de déterminer les fréquences : haute fréquence (> 10 pmo), moyenne fréquence (1-10 pmo), et basse fréquence (< 1 pmo).

2.3 Mesures d'analyses statistiques

Les données ont été analysées avec le logiciel R v4.3 [23] via les : moyennes, écarts-types (ET), médianes, coefficients de variation ($CV = ET/moyenne \times 100$), erreurs absolues (MAE = moyenne $|fre - fa \times 10|$ pour échelle de normalisation) et des quadratiques (RMSE). Il y a aussi une analyse de corrélation linéaire Pearson (r) ; des tests comparatifs : t de Student apparié (fre vs fa normalisée), t indépendant (fre hausse vs fre moyenne) ; puissance statistique post-hoc G*power [24] pour détecter $r > 0.4$ ($1 - \beta = 0.80$) et analyse de régression linéaire simple (fa ~ fre ; fa ~ log (fre)) pour R^2 ajusté. Voir tableau (1) récapitulatif ci-après.

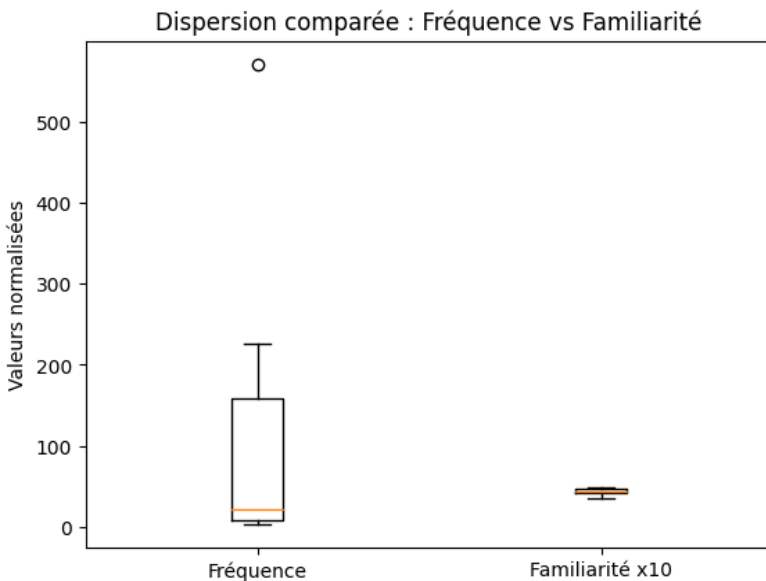
Tableau 1 : Synthèse de la méthode

<i>Type d'analyse</i>	<i>Objectif</i>	<i>Mesures</i>
Analyses descriptives	Quantifier la distribution et la dispersion	CV (ET/moyenne x 100) MAE, RMSE
Tests comparatifs	Détecter les différences systématiques	t Student apparié (fre vs fa normalisée), t indépendant (haute vs moyenne fré)
Corrélation	Mesurer le lien familiarité et fréquence	Pearson : Corrélation linéaire
Puissance statistique	Vérifier la capacité à détecter les effets significatifs	Post-hoc avec G*Power correlations r 0,4, puissance 1- 0 0,80
Régression	Modéliser la relation fréquence et familiarité	Régression linéaire simple : fa fre ; fa log (fre), R ajusté

3. Résultats

Analyses descriptives : Les résultats révèlent une dissociation significative. Sur l'échantillon de 20 items, tous sont jugés familiers ($m_{fa} = 4,36$, $ET = 0,35$, $min = 3,53$, $max = 4,86$) tandis que la fréquence moyenne est de 92.4 pmo ($ET=150,7$; médiane = 18,71 ; $min = 3,34$; $max = 570.3$). L'absolue moyenne (MAE) s'élève à 92.4 pmo et la racine de l'erreur quadratique moyenne (RMSE) à 163.2 pmo, indiquant une variabilité élevée. Le coefficient de variation (CV) souligne l'homogénéité de la fa ($CV_{fa} = 12\%$) contrastant avec la dispersion de fre ($CV_{fre} = 163\%$). On retient de ces analyses descriptives que la dispersion des valeurs de fréquence est nettement plus élevée que celle de la familiarité (normalisée), avec la présence des valeurs extrêmes. À l'inverse, la familiarité présente une distribution resserrée, indiquant une perception globalement homogène des items, malgré des variations importantes de fréquence dans le corpus. La figure (1) ci-après résume la dispersion comparée de fréquence vs familiarité des items.

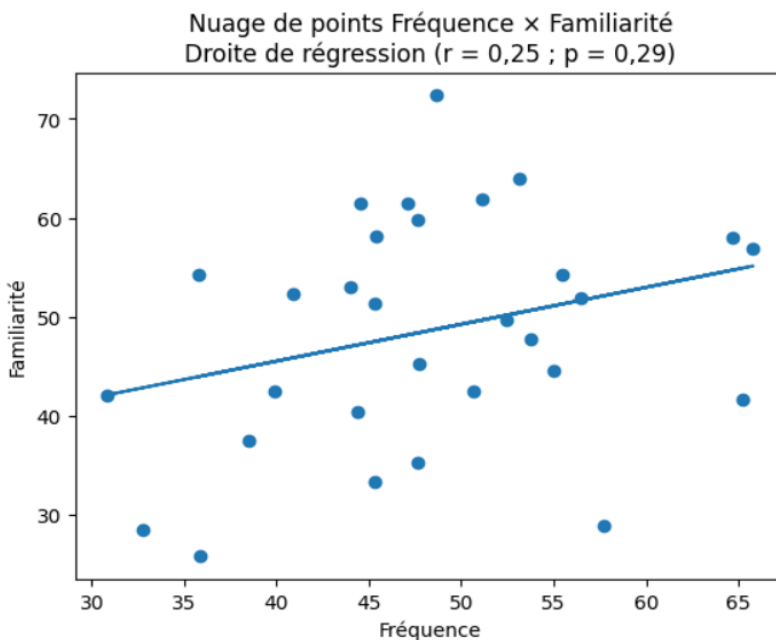
Figure 1 : Dispersion comparée : fa vs fre



Tests comparatifs : On observe des écarts positifs dominants avec $fre > fa$ normalisée, soit 13 items/20, haute fre (Surévalués vs. fa) et 7 items/20, fre moyenne (sous-évalués). On note également qu'il n'y a aucun item de basse fre. Les items de haute fre sont : *maison ; jouer, tomber, lit, dormir ; chien, boire, couper, attraper ; fleur, pomme ; nager, couler*. Et ceux de fre moyenne sont : *ciseaux, chèvre, balai, pleuvoir, banane, scie, cultiver*. Pour le t Student apparié, la comparaison entre les deux variables montre une différence significative de la fre entre items haute fre ($m = 133$, 5 pmo) et moyenne fre ($m = 6.7$ pmo) fréquence ($t(18) = 4,12$, $p < 0,001$, $d \text{ Cohen} = 1,98$), mais non significative pour la familiarité ($m_{haute} = 4.45$ vs $m_{moyenne} = 4,23$; $t(18) = 1.02$, $p = 0,32$). Le test t apparié révèle une tendance à la surreprésentation de fre ($t(19) = 1,84$, $p = 0,08$).

Corrélation : Il existe une faible linéarité entre *fre* et *fa* ($r = 0,25$, $p = 0,29$, IC [95 % 0.22 ; 0.64]). La figure (2) ci-après montre un nuage à points représentant la corrélation. Elle présente un nuage largement dispersé, sans alignement net des points. Une droite de régression légèrement ascendante traduit une tendance positive très faible : lorsque la familiarité augmente, la fréquence tend légèrement à monter de manière instable. Le caractère non significatif du point de vue statistique ($p = 0,29$) confirme l'indépendance de l'une vis-à-vis de l'autre. Malgré l'intervalle de confiance large IC [95 %], l'incertitude reste élevée, ce qui suggère une variabilité inter-items importante. Voir figure (2) pour la corrélation *fre* et *fa* des items lexicaux.

Figure 2 : Corrélation entre *fre* et *fa* des items lexicaux



Puissance statistique : l'analyse Post-hoc avec G*Power révèle un manque de puissance statistique liée à l'échantillon ($n=20$ puissance $< 0,8$ pour $r > 0,4$).

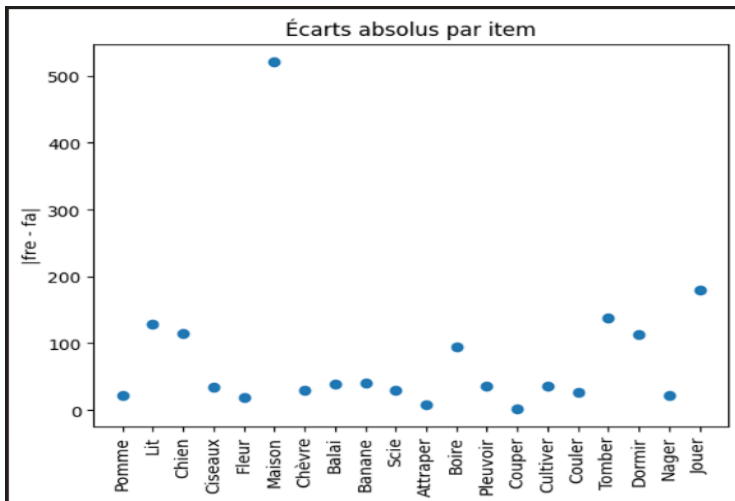
Régression linéaire : Les résultats illustrent la faible linéarité de familiarité. Concrètement, la régression linéaire $fa \sim fre$ produit un faible $R^2 = 0,06$; l'utilisation de $\log(fre)$ n'améliore que marginalement le modèle ($R^2 = 0,12$).

Le Tableau synthétique (2) présente la classification et les écarts par item.

Tableau 2 : Classification et écarts par item

Items	Moyenne (fa)	Fréquence (fre)	Fa x 10	Écart absolu	Écart relatif (%)	Freq. Type
Pomme	4,16	19.77	41,6	-21,83	-52.5	Haute
Lit	4,83	176.1	48,3	+127,8	+265	Haute
Chien	4,36	158.77	43,6	+115,17	+264	Haute
Ciseaux	4,26	7.86	42,6	-34,74	-82	Moyenne
Fleur	4,43	25.2	44,3	-19,1	-43	Haute
Maison	4,86	570.3	48,6	+521,7	+1073	Haute
Chèvre	3,83	8.26	38,3	-30,04	-78	Moyenne
Balai	4,73	8.24	47,3	-39,06	-83	Moyenne
Banane	4,6	6.09	46	-39,91	-87	Moyenne
Scie	3,53	5.21	35,3	-30,09	-85	Moyenne
Attraper	4,26	35.32	42,6	-7,28	-17	Haute
Boire	4,83	142.15	48,3	+93,85	+194	Haute
Pleuvrir	4,36	7.98	43,6	-35,62	-82	Moyenne
Couper	4,4	41.45	43,6	-2,15	-5	Haute
Cultiver	3,86	3.34	38,6	-35,26	-91	Moyenne
Couler	4,16	14.93	41,6	-26,67	-64	Haute
Tomber	4,23	180.25	42,3	+137,95	+326	Haute
Dormir	4,83	160.77	48,3	+112,47	+233	Haute
Nager	4,06	18.71	40,6	-21,89	-54	Haute
Jouer	4,7	225.84	47	+178,84	+380	Haute

Les écarts absolus entre la fréquence et la familiarité normalisée ($fa \times 10$) mettent en évidence une hétérogénéité marquée. Les valeurs observées varient fortement d'un lexème à l'autre avec des écarts particulièrement élevés pour certains items à très forte fréquence (maison, jouer, tomber, dormir), tandis que plusieurs items de fréquence moyenne (banane, balai, chèvre, cultiver) présentent des écarts substantiels. Afin de visualiser la distribution et l'amplitude de ces écarts à un niveau item-spécifique, la figure (3) présente les écarts absolus entre les deux mesures pour l'ensemble des items étudiés.

Figure 3 : Écarts absolus entre fa et fre au niveau item-spécifique

4. Discussion

Cette étude avait pour objectif de mettre en évidence les écarts entre la familiarité lexicale des items en contexte camerounais et leur fréquence en contexte européen francophone - extraite de la base de données standardisées lexique 3 [21]. Cette comparaison entre la familiarité et la fréquence part de la supposition que la fréquence d'un mot est corrélée à sa familiarité, même si la force de cette corrélation n'a pas été étudiée de manière approfondie [25]. En outre, l'approche à petite échelle (20 items adaptés empiriquement en contexte camerounais précisément [7]) permet un diagnostic ciblé et contrôlé, évitant les biais d'un corpus exhaustif tout en révélant des patterns représentatifs d'écarts culturels entre le français européen (lexique 3) et le contexte camerounais francophone.

Les résultats de cette étude diagnostique indiquent une faible association linéaire entre la familiarité subjective des items et la fréquence lexicale objective de ceux-ci ($r \approx 0.25$), alignée sur les observations en psycholinguistique où ces variables, bien que corrélées divergent [20], notamment dans des contextes non standardisés. Cette dissociation est particulièrement prononcée pour les items de moyenne fréquence (ex. : balai, fa = 4,73 malgré fre = 8.24 pmo ; écart rel. 6 7,4 %), reflétant un biais culturel. Cette dispersion inter-items se reflète dans les indices d'erreur globaux (MAE et RMSE), suggérant que les divergences ne sont pas marginales, mais structurantes. Ceci démontre que Lexique 3 – qui est basé sur des corpus littéraires et audiovisuels hexadiques sous-estime les termes ruraux/agricoles centraux au Cameroun (chèvre, banane, cultiver) dont l'usage quotidien élève la familiarité malgré la fréquence modeste. Cette surévaluation subjective de la familiarité par rapport aux fréquences moyennes s'explique par la prédominance de registres oraux ruraux au Cameroun vs corpus basé sur les sous-titres de la télévision et de film comme c'est le cas de Lexique 3 [21]. Elle s'explique aussi par l'influence culturelle, avec des termes tels que « banane », « cultiver » hyperfamiliers localement, contrairement aux corpus occidentaux. Les items de haute fréquence quant à eux (ex. : maison, jouer) montrent des écarts massifs ($+>200\%$), mais masquant une homogénéité de la familiarité ($CV < 15\%$). La faible corrélation entre la familiarité et la fréquence lexicale s'alignent sur les meta-analyses en psycholinguistique démontrant une association modérée globale ($r \approx 0.4-0.6$) qui s'affaiblit dans des contextes non normatifs comme c'est le cas du Cameroun. En effet, cette faible corrélation évoque à nouveau les travaux précédemment cités [20], lesquels montrent que la familiarité subjective des mots (items) explique une part de variation contextuelle – surtout au Cameroun en raison de la sociolinguistique riche complexe de ce pays [5].

Cette dissociation culturelle entre la familiarité des items lexicaux d'intérêt et leur fréquence dans Lexique 3 questionne la pertinence de l'universalité de cette base de données dans l'espace francophone d'Afrique subsaharienne, notamment celle du Cameroun. Ce constat s'accorde à l'affirmation de certains auteurs [26] qui stipulent que le pourcentage de locuteurs d'une population qui connaissent un mot ne corrèle pas parfaitement avec la fréquence telle qu'extraite des corpus classiques (ex. Lexique 3). Ces résultats s'inscrivent donc dans la continuité des travaux psycholinguistiques montrant que la relation entre fréquence et familiarité peut varier selon les corpus étudiés [27] et donc selon les populations. Cette dissociation apparaît particulièrement pour des items de fréquence moyenne (par exemple balai fa = 4,73 malgré fre = 8,24 pmo), ce qui suggère l'existence d'un biais culturel dans la représentation lexicale.

Cette divergence entre fréquence corpus-basée et familiarité peut également être interprétée à la lumière de la forte hétérogénéité sociolinguistique du Cameroun francophone qui produit une large gamme de répertoires langagiers complexes où les locuteurs alternent entre plusieurs codes linguistiques selon les contextes sociaux [5]. Dans ce type d'environnement plurilingue, la familiarité lexicale ne dépend pas uniquement de la fréquence statistique des mots dans les corpus écrits, mais aussi de facteurs socioculturels tels que l'usage oral quotidien, les pratiques professionnelles ou encore les interactions dans les langues locales des différentes et grandes aires culturelles de ce pays [9]. Comme le montrent les travaux sur les normes endogènes, le français camerounais constitue un système linguistique fonctionnel dont les usages sont structurés par des pratiques sociales et des usages endogènes, et non réduits à une simple application de la norme scolaire [12]. Dans cette perspective, la notion de familiarité lexicale reflète moins la fréquence objective dans les corps standardisés que l'exposition cumulative des locuteurs à des contextes d'usage locaux, médiatiques et interactionnels. Cette interprétation est renforcée par l'hétérogénéité du niveau d'éducation, laquelle influence aussi directement l'accès au Français standard – académique. Concrètement, les locuteurs ayant une exposition limitée au Français normatif - appris à l'école - tendent à développer des représentations lexicales ancrées dans les usages endogènes, où les mots les plus fréquents en contexte occidental ne correspondent pas nécessairement aux items les plus familiers cognitivement. Ainsi, les écarts ne doivent pas être interprétés comme de des anomalies, mais comme l'expression d'une réorganisation sociolinguistique des représentations lexicales, structurée par les trajectoires éducatives, les pratiques plurilingues et l'environnement communicatif local. Tenir compte de tous ces facteurs est indispensable dans l'évaluation lexicale en contexte d'aphasie.

En effet, la sélection des items lexicaux pertinents pour développer les batteries d'aphasie est particulièrement cruciale. Les tâches de dénomination d'images, largement utilisées dans cette évaluation reposent généralement sur la familiarité des items pour les locuteurs [28]. Si les items sont culturellement peu familiers, les erreurs observées peuvent refléter un biais culturel plutôt qu'un déficit linguistique réel. Ceci suggère que, dans un contexte comme le Cameroun, l'adaptation des outils d'évaluation devrait reposer sur plusieurs principes : a) la sélection d'items lexicaux familiers pour les locuteurs ; b) l'établissement de normes locales, et c) l'adaptation iconographique des items dans le contexte cible.

Une limite de cette étude est liée à la taille de l'échantillon (20 items lexicaux). Ce nombre restreint manque de puissance statistique ($n=20$ puissance $< 0,8$ pour $r > 0.4$), rendant les écarts observés non généralisables à l'ensemble du lexique. Ce nombre restreint la portée explicative des résultats et impose d'interpréter ces écarts comme relatifs au cadre de comparaison utilisée dans cette étude, plutôt que comme des différences intrinsèques au Français camerounais, limite qui soutient l'argument de [25] selon lequel la corrélation entre ces variables augmente avec la taille des données du corpus. Une autre limite est relative au fait que la familiarité et la fréquence ne sont pas toujours interchangeable comme le soulignent [25]. Bien que le paramètre de familiarité puisse remplacer celui de la fréquence objective dans certaines situations, notamment celles qui impliquent des jugements subjectifs ou des études sur l'expérience individuelle avec le langage [20], ces deux mesures de nature différente – familiarité subjective vs fréquence objective - sont souvent complémentaires et considérées conjointement pour obtenir une vue plus complète des phénomènes linguistiques. Des auteurs [25] affirment d'ailleurs à ce sujet que la familiarité et la fréquence ne sont pas équivalentes même si elles sont corrélées. La non équivalence de ces deux variables, surtout dans le contexte de la présente étude pose un risque de surreprésentation des écarts observés.

En guise de perspectives et comme tenu du caractère exploratoire de cette étude, plusieurs initiatives de recherche peuvent être envisagées à long terme. Premièrement, l'extension du corpus lexical à un nombre plus important d'items permettrait d'améliorer la robustesse statistique des analyses. Deuxièmement, la construction de bases de données psycholinguistiques camerounaises incluant des variables telles que la fréquence, la familiarité, l'imageabilité et l'âge d'acquisition constituerait une ressource essentielle pour la recherche et la clinique. Enfin, des études comparatives impliquant différents profils sociolinguistiques permettraient d'affiner la compréhension des déterminants culturels de la familiarité lexicale dans les contextes francophones africains.

En conclusion, les résultats de cette recherche démontrent qu'il existe des écarts considérables entre la familiarité et la fréquence lexicale des items dans les cadres européen et camerounais francophones. Cette étude soutient donc l'idée que des bases lexicales contextualisées à l'Afrique subsaharienne francophone sont nécessaires pour mieux refléter les usages endogènes qui leur sont spécifiques. Si établies, ces bases de données favoriseront ainsi le développement, la validation et la normalisation d'outils logopédiques et neuropsychologiques adaptés à ces contextes. Celles-ci permettront dans la foulée la vulgarisation de ces domaines dans un contexte où ils sont encore peu implantés. Comme précédemment évoqué, le caractère limité de l'échantillon exploité ne peut prétendre fournir des résultats exhaustifs mais, ceux-ci constituent un bon point de départ pour plaider en faveur de la création desdites bases de données.

Remerciements : Nous exprimons notre gratitude à l'Institut des Sciences Logopédiques de l'Université de Neuchâtel et au Service de Psychologie Cognitive et de Neuropsychologie de l'Université de Mons pour le soutien technique apporté à cette étude.

Financement : Aucun financement spécifique n'a été reçu pour la réalisation de cette étude

Disponibilité des données : les données utilisées dans cette étude sont disponibles auprès de l'auteur de correspondance sur demande.

Contribution des auteurs : Conception de l'étude, réalisation des analyses, interprétation des résultats et rédaction de la première version du manuscrit : N. L., Nguenama. Supervision et correction du manuscrit : L. Lefebvre. Extraction dans les bases de données et correction : S. Invernizzi. Correction du manuscrit : H. B. Edzougou Afanda.

References

1. Organisation Internationale de la Francophonie, Une francophonie de proximité et de terrain : rapport de la secrétaire générale de la francophonie • 2022-2024 https://www.francophonie.org/sites/default/files/2024-10/Rapport%20SG_2022-2024.pdf
2. S. Simeu, Le Français parlé au Cameroun : une analyse de quatre marqueurs discursifs là, par exemple, ékyé et wèé, Thèse de doctorat, Université de Grenoble Alpes, France, 2016
3. S. Kamdem, C. B. O. Ojongnkpot, B. V. Pinxteren, Decolonizing Cameroon's language policies: a critical assessment (2024). *Applied Linguistics Review*. doi:10.1515/applirev-2023-0273
4. G. Piebop, Multilinguisme et dynamique du français au Cameroun : variations, créativité lexico-sémantique et morpho-syntaxique. *FRANCISOLA: Revue de l'enseignement du français et de la littérature en contexte africain et indonésien*. (2016). <https://doi.org/10.17509/francisola.v1i1.2565>
5. G. Echu, The language question in Cameroon. *Linguistik Online*, 18(1), 19–33 (2004) <https://doi.org/10.13092/lo.18.510>
6. E. Chia, J. Suh, A. Ndeffo Tene, Perspectives on Translation and Interpretation in Cameroon (African Books Collective (2009) <https://africanbookscollective.com/publishers/langaa-rpcig-cameroon/>
7. P. Renaud, Le français au Cameroun : fonction (et connotations) d'identité et d'identification ethnique des français régionaux camerounais. *Bulletins des plurilinguismes*. 3, Nice, pp.3-8 (1976)
8. N. L. Nguenama, Une étude d'équivalence linguistico-culturelle préalable au développement d'un test d'aphasie pour la population épileptique adulte du Cameroun francophone, *SHS Web of Conf.*, 191, 10009, (2024).
9. M. D. DeLancey, R. Mbuh & M. W. Delancey *Historical Dictionary of the Republic of Cameroon* 4th ed. Lanham: Scarecrow Press (2010).
10. L.-J. Calvet, *Les voix de la ville : introduction à la sociolinguistique urbaine*. Payot. (1994).
11. R. Chaudenson, *La francophonie : Représentations, réalités, perspectives*. Paris, France : Didier Érudition (1991).
12. C. de Féral, Norme endogène du français au Cameroun, *Collection IDERIC* 12 pp. 65-71 (1991).

13. J. Macoir, M. Fossard, L. Lefebvre, L. Monetta, A. Renard, T. M. Tran, M. A. Wilson, Detection Test for Language Impairments in Adults and the Aged-A New Screening Test for Language Impairment Associated with Neurodegenerative Diseases: Validation and Normative Data. *American journal of Alzheimer's disease and other dementias*, 32(7), 382–392 (2017).
14. J. Macoir, C. Jean, C. Gauthier, Batterie d'Évaluation Cognitive du Langage (BECLA) ». *Journal of Neurological Sciences*, 36, 220-228 (2015).
15. N. L. Nguenama, M. Kamdem, M. Temgoua, C. Ngassam, M. Ngo Ndjeyiha, E. Biloa, L. Lefebvre, M. Fossard, Développement et validation de la Batterie d'Évaluation du Langage chez l'Adulte Aphasique Épileptique (BELAE) du Cameroun francophone *Revue Glossa* (soumis).
16. E. Kaplan, H. Goodglass & S. Weintraub, *The Boston Naming Test*. Lea and Febiger, Philadelphia (1983).
17. A. Kertesz, *Western Aphasia Battery-Revised*. San Antonio, TX: Pearson (2006).
18. S. T. Engelter, M. Gostynski, S. Papa, M. Frei, C. Born, V. Ajdacic-Gross, F. Gutzwiller, & P. A. Lyrer, Epidemiology of aphasia attributable to first ischemic stroke: incidence, severity, fluency, etiology, and thrombolysis. *Stroke*, 37(6), 1379–1384 (2006). <https://doi.org/10.1161/01.STR.0000221815.64093.8c>
19. C. Kuate, B. Tchaleu, M. Motah, S. Nguéfack & J. Doumbe, Situation de l'épilepsie au Cameroun. *North African and Middle East Epilepsy Journal*, 2 (3) (2013).
20. C. M. Connine, J. Mullenix, E. Shernoff, & J. Yelen, Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6), 1084–1096 (1990)
21. B. New, *Lexique 3 : une nouvelle base de données lexicales*, Actes de la 13^e Conférence sur le Traitement Automatique des Langues Naturelles 2 : 892-900 (2006).
22. P. Bonin, R. Peereman, N. Malardier, A. Méot, M. Chalard, A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers* **35**, 158–167 (2003)
23. R Core Team, *R : un langage et un environnement pour le calcul statistique*. R Foundation for Statistical Computing, Vienne, Autriche (2023) <https://www.R-project.org/>

24. F. Faul, E. Erdfelder, A.-G. Lang, A. Buchner, G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191 (2007)
25. K. Tanaka-Ishii, H. Terada, Word familiarity and frequency, *Studia Linguistica*, V. 65 1: 96–116 (2011) <https://doi.org/10.1111/j.1467-9582.2010.01176.x>
26. E. Keuleers, M. Stevens, P. Mandera, M. Brysbaert, Word Knowledge in the Crowd: Measuring Vocabulary Size and Word Prevalence in a Massive Online Experiment. *Quarterly Journal of Experimental Psychology* 68 (8): 1665-1692 (2015)
27. D. A. Balota, M. Pilotti & M. J. Cortese, Subjective frequency estimates for 2,938 monosyllabic words. *Memory & cognition*, 29(4), 639–647 (2001).. <https://doi.org/10.3758/bf03200465>
28. Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 174–215. <https://doi.org/10.1037/0278-7393.6.2.174>