AFEP | Africa Fellows in Education Program

GLOBAL EDUCATION ANALYTICS INSTITUTE
Putting evidence at the heart of education policy

AFEP WORKING PAPER

Linguistic Similarity and Early Literacy Skills: Evidence from Cameroon

Flora Yselle Malah Kuete

University of Yaoundé II, Cameroon

May 2025

Acknowledgements

This work was carried out with financial and scientific support from the Partnership for Economic Policy (PEP) and the Global Education Analytics Institute (GEAI) through funding provided by the Yidan Foundation.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Partnership for Economic Policy (PEP) or the Global Education Analytics Institute (GEAI).

Linguistic Similarity and Early Literacy Skills: Evidence from Cameroon

Flora Yselle Malah Kuete*

University of Yaoundé II, Cameroon

May 2025

Abstract

Over the past decades, early childhood education policies have shifted focus from access to learning quality, acknowledging that school attendance is insufficient when children do not understand the language of instruction (LOI). This paper examines the impact of linguistic similarity between students' home languages (L1) and the LOI on early literacy skills in Cameroon, a multilingual country with over 270 living languages. Using the 2019 Early Grade Reading Assessment data and lexicostatistics, I construct a quantitative measure of phonetic similarity between 15 major indigenous languages and the two official LOIs, French and English. I then employ the two-stage least squares method to examine the effect of linguistic similarity on foundational skills. The results show that greater phonetic similarity significantly improves listening comprehension, particularly within the Francophone education system. The effects are also more pronounced among students who receive reading support from parents or siblings, suggesting that home literacy environments reinforce the benefits of linguistic proximity. Additionally, phonetic similarity appears to mitigate learning disadvantages in less privileged regions. These findings underscore the importance of integrating linguistic alignment into language-in-education policies to strengthen early learning outcomes and advance equity in multilingual contexts.

Keywords: mother tongue; language of instruction; phonetic similarity; listening comprehension; Cameroon. *JEL codes:* I21, I24, O15, H52, H75

* Faculty of Economics and Management, University of Yaoundé II, Cameroon. Email: florayselle27@yahoo.com

1 Introduction

Improving the quality of education is at the forefront of global development agendas, shifting the focus from access to learning outcomes (World Bank, 2018). Although school enrollment rates have increased across sub-Saharan Africa, millions of children still complete primary education without acquiring basic literacy skills (PASEC, 2019). Among the persistent challenges, the language of instruction (LOI) remains a critical barrier. In many multilingual societies, children enter school without prior exposure to the LOI, creating a fundamental disconnect between the language of schooling and the language of the home. While Sustainable Development Goal 4 (SDG 4) promotes inclusive and equitable education, it does not sufficiently address the role of language in shaping learning trajectories in linguistically diverse contexts (UNESCO, 2022).

Early theoretical models have shaped our understanding of language transfer in education. The Interdependence Hypothesis posits that cognitive and academic skills acquired in a learner's first language can support learning in a second language (Cummins, 2000). Similarly, the Lexical Quality Hypothesis suggests that the quality of word representations in long-term memory is stronger when formed in a familiar language and is essential for reading comprehension (Perfetti and Hart, 2002). These frameworks highlight the cognitive advantages of using the home language (L1) during the initial stages of schooling.

However, studies that test these hypotheses are geographically and linguistically skewed. A systematic review by Bylund et al. (2023), which analyzed over 2,000 articles published in major journals on second language acquisition and multilingualism, found that the majority of studies were conducted in North America and Western Europe. Highly multilingual regions in the Global South, such as sub-Saharan Africa, were largely absent from the evidence base.

Nonetheless, some empirical studies in Africa have confirmed the benefits of mother-tongue instruction. For example, research in Mozambique, Mali, Nigeria, Malawi, and Kenya consistently shows that children taught in their mother tongue achieve better reading, math, and school attendance outcomes (Benson, 2004; Fafunwa et al., 1989; Chimombo, 2009; Piper et al., 2016).

However, most of this literature focuses on the extensive margin through a dichotomous lens, classifying children as matched or mismatched with the LOI. It overlooks that some indigenous languages are structurally or phonetically closer to French or English than others. A learner whose L1 shares many phonetic features with French may experience fewer difficulties than a peer whose L1 is linguistically distant. This gap in the literature highlights the need for more granular, continuous measures of linguistic similarity that capture the intensive margin of language alignment.

A growing number of studies are attempting to move beyond this dichotomy by introducing the concept of linguistic distance. For instance, Chiswick and Miller (2008) developed a linguistic distance index to assess English proficiency among immigrants in the U.S. and Canada,

finding that greater distance correlated with lower proficiency. More recent contributions, such as Igarashi et al. (2024), applied this approach to the Philippines, where learners whose L1s were closer to the LOI performed better in reading and math, regardless of socioeconomic status. However, such approaches remain rare in the Sub-Saharan Africa context, and they often rely on lexical similarity or genealogical proximity rather than phonetic alignment, which can be misleading in highly multilingual societies.

This study addresses these limitations by developing a phonetic similarity index that captures the intensive margin of language alignment. The index is based on a lexicostatistical framework comparing 15 major indigenous Cameroonian languages, self-reported by students as their L1s, with the two official languages of instruction: French and English. Cameroon presents an ideal context for this analysis. With over 270 spoken languages, the country operates two distinct education systems, Francophone and Anglophone, where French and English serve as the LOIs. Yet, most children speak neither language at home, creating a linguistic mismatch between school and home. Unlike other African countries that have adopted bilingual or mother-tongue-based instruction in early grades, Cameroon has maintained French and English as the sole LOIs.

Existing studies in Cameroon underscore both the promise and limitations of incorporating local languages into formal education. Programs such as PROPELCA and the Kom Experimental Project have shown that instruction in indigenous languages can enhance student engagement and academic performance (Trudell, 2005; Laitin et al., 2019). Comparative research between the Francophone and Anglophone systems has highlighted structural and pedagogical differences, with some evidence suggesting better outcomes among Francophone students (Nana, 2013; Bekkouche and Dupraz, 2023). Further studies confirm that students tend to perform better when the LOI aligns with the language spoken at home (Avom et al., 2021; Kouamo and Dieng, 2022; PASEC, 2019). However, this body of work relies primarily on an extensive approach, overlooking the potential significance of partial or graded linguistic alignment.

Using data from 459 third-grade students in the 2019 Early Grade Reading Assessment (EGRA), this paper investigates whether greater phonetic similarity between a child's L1 and the LOI enhances listening comprehension, a fundamental skill for literacy acquisition. To mitigate potential endogeneity, where unobserved factors might influence both language exposure and literacy outcomes, the empirical analysis employs a two-stage least squares (2SLS) instrumental variable strategy. The instruments are based on the geographic distances between each student's province and major cities where French or English is widely spoken: Yaoundé and Douala for French; Buea and Bamenda for English. The identification strategy assumes that linguistic diffusion is influenced by geographic proximity, but that distance does not directly affect literacy outcomes.

The results indicate that higher phonetic similarity between L1 and the LOI significantly enhances listening comprehension in early primary grades. Specifically, a one-unit increase

in phonetic similarity with French is linked to a statistically significant improvement of 0.012 standard deviations in listening comprehension scores, and 0.008 for similarity with English. However, these results are not consistent across all learners. Subgroup analysis reveals that the effect is stronger in the Francophone system and in rural areas, where children's exposure to the LOI outside the classroom is limited. In such contexts, linguistic proximity seems to facilitate the transition into formal schooling. Additionally, home literacy environments also influence the relationship. Children who engage in reading at home or receive homework help from parents or siblings benefit more from phonetic alignment. In contrast, phonetic similarity alone does not significantly improve outcomes for children with minimal exposure to print or language support at home.

This study contributes to the literature in three significant ways. First, it reconceptualizes linguistic alignment as a continuum, introducing a phonetic similarity index that captures intensive margin variation. Second, it proposes a scalable methodology rooted in articulatory phonetics rather than lexical or genealogical proximity, making it better suited to contexts where the L1 and LOI do not belong to the same language family. Third, it applies this index within a multivariate framework that accounts for heterogeneity in education system type, locality, and household learning conditions, offering a more nuanced understanding of how language alignment shapes early literacy.

The remainder of the paper is structured as follows. Section 2 describes the Cameroonian education system and linguistic context. Section 3 outlines the data and measurement strategies. Section 4 presents the empirical methodology. Section 5 discusses the main results, and Section 6 concludes with policy implications and directions for future research.

2 Background

Cameroon offers a distinctive setting where two formal education systems function within one of the world's most linguistically diverse nations. To understand how language policy impacts learning outcomes, it is essential to first examine the structure of the education system and then explore the broader multilingual environment in which children develop literacy.

2.1 Overview of Cameroon's Dual Education System

Cameroon's education system is officially bilingual and organized into two distinct yet coexisting subsystems: Francophone and Anglophone. This duality reflects the country's colonial past, wherein French and British administrations implemented separate education models in their territories. After reunification in 1961, both systems were preserved and incorporated into a national framework governed by the Ministère de l'Éducation de Base (MINEDUB) for pre-

primary and primary levels, and the Ministère des Enseignements Secondaires (MINESEC) for secondary education. The system is governed by Law No. 98/004 of April 14, 1998, which guarantees equal access to quality education for all citizens and sets the foundation for education governance in Cameroon.

Children typically begin school at age six, following an optional pre-primary phase for children aged three to five. Education is compulsory up to age 11, ensuring that all children receive basic education regardless of socioeconomic background.

The Francophone subsystem, dominant in eight out of ten regions, is modeled after the French system. It includes six years of primary education culminating in the Certificat d'Études Primaires (CEP), followed by a secondary cycle comprising four years for the first cycle (ending with the Brevet d'Études du Premier Cycle, BEPC) and three years for the second cycle, which concludes with the Baccalauréat. Emphasis is placed on memorization, grammar, and structured assessments, with a curriculum anchored in mathematics, French literature, and national history. The Anglophone subsystem, prevalent in the North-West and South-West regions, mirrors the British system. It consists of a six-year primary cycle ending with the First School Leaving Certificate (FSLC), followed by a five-year lower secondary cycle leading to the General Certificate of Education Ordinary Level (GCE O-Level), and a two-year upper secondary cycle culminating in the GCE Advanced Level (A-Level). Pedagogically, it promotes interactive learning, critical thinking, and subject flexibility, particularly in English literature, sciences, and social studies.

In addition to curricular and pedagogical differences, cultural perceptions differentiate the two systems. The Francophone model is widely viewed as academically rigorous and formal, while the Anglophone system is considered more dynamic and student-centered. These perceptions influence parental preferences and school selection across regions. While regionalized in origin, both subsystems are found nationwide. Migration, perceived academic advantages, and the socioeconomic benefits of bilingualism have led to a nationwide dispersion of both models. This bilingual approach gives learners broader prospects at both national and international levels.

Cameroon's schools are also categorized as secular or faith-based. Secular schools are government-run and offer religiously neutral curricula. Faith-based schools, typically affiliated with Christian or Islamic institutions, integrate moral and religious instruction alongside standard academics. Despite differing institutional values, both types lead to the same public examinations.

Educational access has steadily improved over the past three decades. The net enrollment rate in primary school increased from 73% in 1995 to over 91% by 2018, while the gross enrollment ratio reached 112% by 2022 (World Bank, 2023). Gender parity has considerably improved in the early grades. Yet, quality remains a challenge. The 2019 PASEC assessment revealed that 60.6% of Grade 2 students failed to reach basic literacy proficiency, with 11.8% scoring below the

minimum threshold (CONFEMEN, 2020). In 2018, the primary repetition rate stood at 23.5%, placing Cameroon among the highest in Sub-Saharan Africa (World Bank, 2020). Dropout rates surge at the transition to secondary education, particularly in linguistically marginalized communities. Girls face additional obstacles related to early marriage, social norms, and the lack of gender-responsive environments.

Several systemic issues hinder progress. The education sector is underfunded, leading to inadequate infrastructure, overcrowded classrooms, and shortages of learning materials. Geographic disparities persist, particularly in rural and underserved areas, where schools often lack qualified staff and basic amenities. Linguistic factors further compound these structural inequalities. Most children enter school speaking a local language and must quickly transition to French or English. This linguistic leap is particularly burdensome in rural communities, where exposure to the LOI is minimal and support at home is limited due to parents' low proficiency.

Cameroon has aligned its higher education with global standards. Its participation in the Bologna Process facilitates academic recognition across Europe. The Baccalauréat and GCE A-Level enable students to access universities in Francophone and Anglophone countries, respectively, supporting academic mobility and regional integration.

2.2 Linguistic Diversity and Language of Instruction

With 273 living languages, Cameroon ranks as the second most linguistically diverse country in Africa after Nigeria and the ninth globally (Eberhard et al., 2024). These languages belong to four major language families: Niger-Congo, Bantu, Chadic, and Nilo-Saharan. Multilingualism is common across regions, with most children growing up speaking more than one local language. However, this diversity is not reflected in national education policy.

French and English are the only official languages of instruction in public schools. This creates a disconnect between the languages children speak at home and those used in the classroom. Many pupils begin school without any prior exposure to the LOI. This gap is especially significant in rural areas, where families often do not speak the LOI and are unable to support early literacy. Teachers receive limited training in multilingual pedagogy, and learning materials are almost exclusively available in French or English (Benson, 2005; Ball, 2010).

Some localized initiatives have sought to address this challenge. The Kom Experimental Mother Tongue Education Project, for instance, introduced instruction in Kom, the local language of the Boyo division, during the first three years of primary school. This initiative led to improvements in reading fluency, pupil retention, and parental involvement (Chiatoh, 2012). Another example is the PROPELCA program (Projet de Recherche Opérationnelle pour l'Enseignement des Langues Camerounaises). This initiative developed orthographies, primers, and teacher training materials for indigenous languages such as Bafut, Kom, and Nso (Tadadjeu and Sadembouo, 2000).

Despite their promising outcomes, these programs remain limited in scale. They have not been integrated into the national education strategy due to political sensitivities surrounding language policy, insufficient funding, and a lack of institutional coordination (Trudell, 2009).

In contrast, several Sub-Saharan African countries have formally incorporated indigenous languages into the curriculum. For instance, Ethiopia allows instruction in languages like Afaan Oromo, Amharic, and Tigrinya through primary school, with English introduced progressively in later years (Heugh et al., 2007). Malawi promotes the use of Chichewa and other local languages from Grades 1 to 4 (Chimombo, 2009). In Rwanda, children begin schooling in Kinyarwanda before transitioning to English in the upper primary grades (Williams and Ball, 2014). These experiences have been linked to improvements in literacy, school participation, and learning equity (Piper et al., 2016).

Cameroon's ongoing reliance on French and English as the sole languages of instruction in the early years raises important questions about the inclusiveness and effectiveness of its education system. This situation highlights the broader debate on the role of language in shaping foundational learning. The following section examines the evidence from empirical research on how the language of instruction influences early educational outcomes, particularly in multilingual contexts.

3 Data

This study utilizes data from the 2019 Early Grade Reading Assessment (EGRA) in Cameroon, an evaluation designed to measure foundational literacy skills among young learners. It is administered by the Ministry of Basic Education and the Ministry of Secondary Education under the *Programme d'Amélioration de l'Équité et de la Qualité de l'Éducation au Cameroun (PAEQUE)*. The evaluation adheres to international standards to ensure robust and comparable results. Data collection employs electronic tablets and the *Tangerine* platform, which streamlines data management and reduces entry errors. In addition to literacy tests, contextual questionnaires collect information from students, teachers, and school administrators. These questionnaires capture demographic characteristics, home language use, parental literacy levels, and access to educational resources.

I selected EGRA over other assessments, such as PASEC, because it uniquely collects information on students' home languages. One key question asks students if their LOI matches their L1. If it does not, they specify their home language, enabling the identification of linguistic mismatches, a central variable in this study.

The EGRA assessment occurred in two phases during the 2018–2019 academic year. The first phase, in January 2019, assessed foundational literacy skills expected by the end of Grade 2. The second phase, in May 2019, evaluated skills students acquire by the end of Grade 3. To

ensure representative results, EGRA employed a stratified cluster sampling method that covered both urban and rural schools. The final sample included 80 public schools, equally distributed between Francophone and Anglophone subsystems, with approximately 800 students assessed in the first phase and 960 in the second. This study focuses exclusively on Phase 2 data because Phase 1 lacks sufficient information on students' home languages to generate a reliable sample size.

After processing the responses, students report speaking 29 indigenous languages alongside French and English. However, this study includes only 15 indigenous languages, selected based on their recognition as written languages in Ethnologue. By filtering for complete home language data, the study focuses on a subset of 459 students across eight of Cameroon's ten regions. Due to security concerns, the assessment excludes the North-West and South-West regions, where ongoing conflicts disrupted data collection efforts.

3.1 *Literacy skill*

The outcome variable in this study is literacy skills, measured through listening comprehension. In educational research, listening comprehension is a widely recognized measure of literacy skills. For instance, Kim and Phillips (2014) demonstrated that listening comprehension significantly predicts reading comprehension outcomes in early elementary students. Similarly, Hogan et al. (2014) found that listening comprehension is dominant in reading comprehension as early as the elementary grades. The EGRA assessment evaluates whether students' difficulties in other literacy-related tasks such as recognizing alphabetic principles, reading familiar words, decoding unfamiliar words, or understanding full texts originate from decoding issues or other linguistic barriers. The test administrator reads a short passage aloud, followed by five comprehension questions to assess the student's oral understanding.

The five questions[1] target different levels of comprehension: four assess contextual understanding, and one requires making an inference. Responses are recorded in a binary format, with 1 assigned to correct answers and 0 to incorrect ones. To facilitate interpretation in econometric models, I construct a continuous variable by dividing the total number of correct responses by 5, the total number of items. This yields a fractional score ranging from 0 to 1, representing the proportion of correct answers. This scoring method aligns with international standards used in large-scale assessments and offers greater flexibility for analysis within linear and limited dependent variable frameworks.

[1] The questions are as follows: "Where did the little hen fall?" (listening_comprehension_1), "What color is the goat?" (listening_comprehension_2), "Who fell last into the pond?" (listening_comprehension_3), "What object did the little hen see to save them?" (listening_comprehension_4), and "Which of the two animals, the hen or the goat, came up with the idea that saved them?" (listening_comprehension_5).

3.1.1 Linguistic Similarity

A key contribution of this study is introducing a quantitative measure of linguistic similarity between the LOI and students' L1 in Cameroon. Empirical studies have developed various methodologies to quantify linguistic distance, demonstrating its impact on language acquisition, cognitive processing, and educational outcomes. For instance, Chiswick and Miller (2005) measured linguistic distance by comparing structural, phonological, and lexical differences between English and other languages. Their findings showed that greater linguistic distance negatively affects language acquisition and economic integration among immigrants in the United States and Canada. Serva and Petroni (2008) proposed an orthographic distance measure, which quantifies spelling differences between words in various Indo-European languages. Their computational approach reconstructed linguistic trees and objectively classified language similarity.

Other studies have utilised lexical similarity as a proxy for linguistic distance. Wichmann et al. (2010) applied the Automated Similarity Judgment Program (ASJP) to analyze phonetic representations of core vocabulary words across various languages. Their approach systematically classified language families and computationally estimated linguistic proximity. More recently, Igarashi et al. (2024) introduced a linguistic distance measure to evaluate the effects of mother tongue-based education in the Philippines. Their approach follows Fearon (2003) and Nakagawa and Sugasawa (2022), defining linguistic similarity as the number of shared edges between two languages in a linguistic tree, normalized by the maximum possible shared edges. De Gregorio et al. (2024) extended linguistic distance research by analyzing syntactic differences between languages, using part-of-speech distributions to quantify syntactic distance. Their findings revealed that syntactic similarity closely correlates with mutual intelligibility and plays a critical role in language learning.

Building on these studies, this research adopts a lexicostatistical framework to construct a phonetic similarity index between students' L1 and the LOI. Unlike lexical or syntactic criteria, the phonetic criterion focuses on how closely the sounds of words align in spoken language, which is crucial for comprehension and literacy acquisition (Schepens et al., 2020; Booth et al., 2006). This method allows for comparisons across diverse language families without requiring a common linguistic origin. Furthermore, while lexical similarity relies on shared etymology and cognate identification, phonetic similarity enables comparisons between languages that do not share a common linguistic ancestry (McMahon and McMahon, 2005). This is particularly valuable in Cameroon, where L1s belong to multiple language families, including Niger-Congo (e.g., Bantu languages such as Ewondo and Bulu), Nilo-Saharan (e.g., Fufulde and Dii), and Afro-Asiatic (e.g., Mafa and Guidar).

To quantify phonetic similarity, the study employs a standardized list of 164 core words derived from the *Atlas Linguistique du Cameroun* (Dieu and Renaud, 1983). National linguists have phonetically transcribed these words for each of the 15 L1s included in the analysis: Pid-

gin English, Yemba, Nda'Nda, Fe'efe, Kako, Bamun, Bulu, Ewondo, Fufulde, Guidar, Mafa, Dii, Doyayo, Daba, and Basaa, along with the two LOIs, French and English (see Table A1 in the appendix for complete transcriptions). These transcriptions serve as a basis for pairwise comparisons conducted using the Cog Language Comparison Software.

The Cog software applies a phonetic alignment algorithm to compute linguistic similarity between each L1 and the LOI. Word pairs that exhibit phonetic similarity above a predefined threshold are identified as cognates, ensuring an objective and systematic classification. This approach follows principles used in automated cognate detection and sound alignment, as described in List (2012, 2014), and is particularly suited for large-scale cross-linguistic comparisons where manual phonological comparison is infeasible. The software calculates a phonetic similarity index for each L1-LOI pair, capturing linguistic proximity on a continuous scale rather than a binary classification. The phonetic similarity percentage (S_{ij}) between L1 i and LOI j is computed using the following formula:

$$S_{ij} = \left(\frac{Number\ of\ similar\ phonemes\ across\ words\ in\ i\ and\ j}{Total\ number\ of\ phonemes\ compared}\right) \times 100 \quad (1)$$

For each word, the method identifies phonetic matches based on articulation features such as voicing, manner of articulation, and place of articulation. The total number of matching phonemes is then divided by the total number of phonemes compared across all 164 words. This calculation produces a percentage representing phonetic alignment between an L1 and the LOI. Students who report French or English as their L1 automatically receive a similarity score of 100% when the LOI matches their L1.

Table 1 presents the results of the phonetic similarity matrix, which compares the listed indigenous languages to English and French. The table reveals that Pidgin English exhibits the highest phonetic similarity with English (78%) and a notable similarity with French (35%). This result aligns with expectations, given Pidgin English's origins as an English-based creole. Beyond Pidgin English, languages such as Kako (41%) and Daba (40%) also show relatively high phonetic similarity to English. On the French side, Kako (56%) and Nda'Nda (50%) stand out as the most phonetically similar languages. Interestingly, the similarity between the two official languages, English and French, is 39%. While this figure is moderate, it exceeds the similarity levels observed between most indigenous languages and the official LOIs. These results underscore the linguistic diversity in Cameroon and highlight the varying degrees of phonetic alignment between local languages and the LOI, which may have implications for literacy acquisition and learning outcomes.

Figure A2 in the appendix presents the phonetic dendrogram and network graph, which visually depict the relationships between languages. The dendrogram clusters languages based on their phonetic similarity, where shorter branch lengths indicate stronger linguistic proximity. As expected, Pidgin English and English form a close cluster, reflecting their high phonetic simi-

Table 1: Linguistic Similarity Matrix

	English	*French*	*Pidgin*	*Yemba*	*Nda'Nda*	*Fe'efe*	*Kako*	*Bamun*	*Bulu*	*Ewondo*	*Fufulde*	*Guidar*	*Mafa*	*Dii*	*Doyayo*	*Daba*	*Basaa*
English		**39**	**78**	**26**	**18**	**26**	**41**	**30**	**26**	**23**	**29**	**28**	**32**	**39**	**28**	**40**	**35**
French	**39**		**35**	**27**	**50**	**25**	**56**	**33**	**30**	**25**	**33**	**38**	**33**	**29**	**33**	**24**	**25**
Pidgin Eng.	**78**	**35**		26	18	24	41	29	29	29	27	32	30	39	26	38	33
Yemba	**26**	**27**	26		69	52	40	28	27	33	32	32	32	20	30	32	36
Nda'Nda	**18**	**50**	18	69		71	0	0	22	22	14	0	0	0	28	0	0
fe'efe	**26**	**25**	24	52	71		50	36	38	40	23	19	28	28	20	0	36
Kako	**41**	**56**	41	40	0	50		79	0	0	43	0	39	37	35	0	0
Bamun	**30**	**33**	29	28	0	36	79		32	33	33	19	34	29	35	19	31
Bulu	**26**	**30**	29	27	22	38	0	32		88	28	24	28	32	29	27	47
Ewondo	**23**	**25**	29	33	22	40	0	33	88		29	0	33	19	29	21	40
Fufulde	**29**	**33**	27	32	14	23	43	33	28	29		50	35	31	36	34	31
Guidar	**28**	**38**	32	32	0	19	0	19	24	0	50		55	21	49	0	33
Mafa	**32**	**33**	30	32	0	28	39	34	28	33	35	55		28	32	41	34e
Dii	**39**	**29**	39	20	0	28	37	29	32	19	31	21	28		49	32	27
Doyayo	**28**	**33**	26	30	28	20	35	35	29	29	36	49	32	49		32	32
Daba	**40**	**24**	38	32	0	0	0	19	27	21	34	0	41	32	32		15
Basaa	**35**	**25**	33	36	0	36	0	31	47	40	31	33	34	27	32	15	

Notes: **Source**: Author estimates using the Cog Phonetic Method Settings. **Note**: The values represent the phonetic similarity percentage between language pairs, calculated by comparing phonetic transcriptions of 164 standardized words. For example, a value of 78 between Pidgin English and English means 78% of the phonemes are phonetically similar, indicating closer linguistic alignment.

larity consistent with the matrix data. Similarly, in the network graph, Pidgin English, English, Bamun, Kako, and French exhibit strong connections, represented by thicker lines. In contrast, most indigenous languages, such as Ewondo, display a more distant and dispersed relationship with the official LOIs, as illustrated by thinner and more widely spread connections.

3.1.2 Based on these results, the study constructs two measures of linguistic similarity: one capturing the similarity between each indigenous language and English, and the other capturing the similarity with French. The values for these measures correspond to those presented in Table 1. For example, a student whose L1 is Nda'Nda receives a score of 50 for the "similarity with French" variable and 19 for the "similarity with English" variable. The same coding applies to other languages.

3.1.3 Summary statistics

Table 2 presents descriptive statistics for all variables used in the analysis. The listening comprehension score, which captures the mean of correct answers across the five oral comprehension questions, has a mean of 0.321 and a standard deviation of 0.345. The relatively low mean highlights modest literacy performance across the sample, while the wide dispersion indicates substantial variation in comprehension skills among students.

Phonetic similarity between students' mother tongues and the two official languages reveals substantial variation across the sample. On average, students' L1s are more phonetically similar to English than to French, suggesting that, in general, English may be linguistically more accessible for a larger share of learners. The average similarity score with English is approximately 59, compared to 42 for French, indicating a broader baseline alignment with English. However, the large standard deviation for English (34) relative to that of French (20) points to greater dispersion, some students speak languages that are extremely close to English, while others are very distant from it. The range of similarity scores (from 18 to 100 for English, and from 24 to 100 for French) further underscores the extent of linguistic heterogeneity. While a subset of students report an official language as their L1 (similarity score of 100), others face considerable phonetic distance from the LOI.

Demographically, the average student is 8.8 years old, aged 6 to 14 years. Gender is nearly evenly distributed, with girls representing 51.6% of the sample. Parental literacy remains relatively low: only 40.7% of fathers and 38.3% of mothers are reported as able to read. Household asset indicators, used as proxies for socioeconomic status, show moderate access to essential utilities and educational resources. Approximately 50.5% of students have access to a radio, 58.0% to electricity, and 51.2% to a television. In contrast, only 17.6% of students report having access to a computer, reflecting limited exposure to digital learning tools. Regarding educational environments, about 61.2% of the sample are enrolled in the Francophone school system, consistent with national demographic trends. Statistics also report that approximately 33.1% of students in

Table 2: Summary Statistics of Study Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Listening comprehension score	459	0.321	0.345	0	1
Similarity with French	459	41.682	20.350	24	100
Similarity with English	459	58.983	33.831	18	100
Age	459	8.826	1.335	6	14
Sex (Female = 1)	459	0.516	0.500	0	1
Father is literate	459	0.407	0.492	0	1
Mother is literate	459	0.383	0.487	0	1
Radio at home	459	0.505	0.501	0	1
Electricity at home	459	0.580	0.494	0	1
TV at home	459	0.512	0.500	0	1
Computer at home	459	0.176	0.382	0	1
Francophone system (1=yes)	459	0.612	0.488	0	1
Locality (1 = Rural)	459	0.331	0.471	0	1
Reads at home	454	0.764	0.425	0	1
Parents help with homework	458	0.260	0.439	0	1
Siblings help with homework	458	0.172	0.378	0	1
No one helps with homework	458	0.207	0.406	0	1
Log distance to Buea	459	6.025	0.901	4.315	7.260
Log distance to Bamenda	459	5.896	0.820	3.918	7.041
Log distance to Yaoundé	459	6.185	0.742	4.397	7.238
Log distance to Douala	459	6.192	0.804	4.103	7.305
Log distance to Ebolowa	459	6.290	0.757	4.533	7.340

Notes: Source: Author's calculations.

the sample are from rural areas. This distribution underscores the urban bias of the sample, but still allows for meaningful comparisons. Additional variables capturing learning support at home show that 76.4% of students report reading outside of school, while 26.0% receive homework help from parents, 17.2% from siblings, and 20.7% from no one. These indicators highlight the uneven availability of academic support in the home environment. Finally, geographic controls reflect the natural log of road distances from students' schools to key urban centers. Mean log distances range from 5.896 to 6.290, with variation across schools reflecting the diversity of the sample's geographic distribution.

4 Empirical Strategy

Estimating the causal effect of linguistic similarity on literacy outcomes raises endogeneity concerns. Linguistic similarity between a student's L1 and the LOI may be correlated with unobserved determinants of learning, such as household socioeconomic status, parental education, school quality, or historical migration patterns. Ignoring these factors could bias the estimated relationship between phonetic proximity and literacy outcomes.

Formally, the underlying education production function can be expressed as:

$$y_i = \alpha + \rho Sim_i + X_i'\beta + \epsilon_i \tag{2}$$

Where y_i is the listening comprehension outcome for student i, Sim_i is the linguistic similarity between the student's L1 and the LOI, X_i' is a vector of observed control variables (age, sex, father and mother can read, get radio, electricity, TV and computer at home), and ϵ_i is an error term. The coefficient ρ is the parameter of interest and captures the effect of linguistic similarity on comprehension. This model can be estimated via OLS, but if Sim_i is endogenous, due to omitted variables or reverse causality, OLS estimates of ρ will be biased and inconsistent.

To address this concern, I use a two-stage least squares (2SLS) strategy with instrumental variables (IV). A valid instrument must meet the condition of relevance by being sufficiently correlated with the endogenous regressor and the condition of exogeneity by influencing the outcome solely through its effect on that regressor, without any direct impact on the dependent variable.

I use distances from each student's province to the cities where the LOI is predominantly spoken as instruments. These distances reflect long-run linguistic exposure but are unlikely to directly affect listening comprehension. The intuition is that proximity to cities where the LOI is historically dominant increases the likelihood of phonetic convergence between the local language and the LOI due to linguistic diffusion. This spatial IV approach is inspired by the logic of using geographic proximity as a source of exogenous variation in individual exposure, similar in spirit to Card (1993), who uses distance to college as an instrument for education.

The strategy builds on the notion that language similarity follows spatial rather than purely individual patterns, as highlighted by Bosker and Buringh (2017) and Michalopoulos (2012). The core identifying assumption is that geographic distance affects phonetic similarity due to long-term linguistic contact but does not influence literacy outcomes directly, except through this channel. For example, the Northwest region, which borders Nigeria, shares greater phonetic proximity with English due to cross-border linguistic contact (Weiss, 1996). Conversely, more isolated Francophone provinces retain greater linguistic distance from French. Since listening comprehension is primarily influenced by individual cognitive development (Snow, 2010), school quality (Glewwe and Muralidharan, 2016), and home literacy environments (Kim and

Phillips, 2014), it is unlikely that geographic distance per se directly affects learning, satisfying the exclusion restriction.

To construct the geographic distance variables, the 2018 Cameroon Demographic and Health Survey data is used to identify regions and cities with the highest concentrations of French and English speakers (see Figure A3 in the Appendix). The results indicate that Buea and Bamenda have the highest proportion of English speakers, while Yaoundé and Douala have the highest proportion of French speakers. Once these cities are identified, four geographic distances between the centroid of each identified LOI-majority city are computed (Buea and Bamenda for English; Yaoundé and Douala for French) and the centroid of the capital city of the student's school province, as reported in the EGRA. This process yields four instrumental variables, two for English and two for French[2].

The empirical model follows a standard 2SLS framework. In the first stage, linguistic similarity is regressed on the instruments and control variables:

$$Sim_i = \alpha + Z_i' \beta + X_i' \gamma + \epsilon_i \tag{3}$$

where Z_i' is the vector of instruments and X_i' includes controls for student characteristics, parental literacy, household assets (electricity, TV, computer), and region fixed effects. In the second stage, the predicted similarity is used to estimate the effect on listening comprehension:

$$Comp_i = \theta + \delta.\widehat{Sim_i} + X_i' \gamma + u_i \tag{4}$$

where $Comp_i$ is the listening comprehension score for student i, and $\widehat{Sim_i}$ is his predicted phonetic similarity obtained from the first stage. X_i represents the same set of control variables. The coefficient δ captures the causal effect of linguistic similarity on the listening comprehension test score. The validity of the instruments is empirically confirmed through over-identification checks.

5 Results

This section presents the results in four parts: descriptive evidence on geographic distribution concerning linguistic similarity and listening comprehension; econometric results on the causal effect of linguistic similarity; heterogeneity analyses by school system, localities, and home learning conditions; and an exploration of whether similarity mitigates regional disparities.

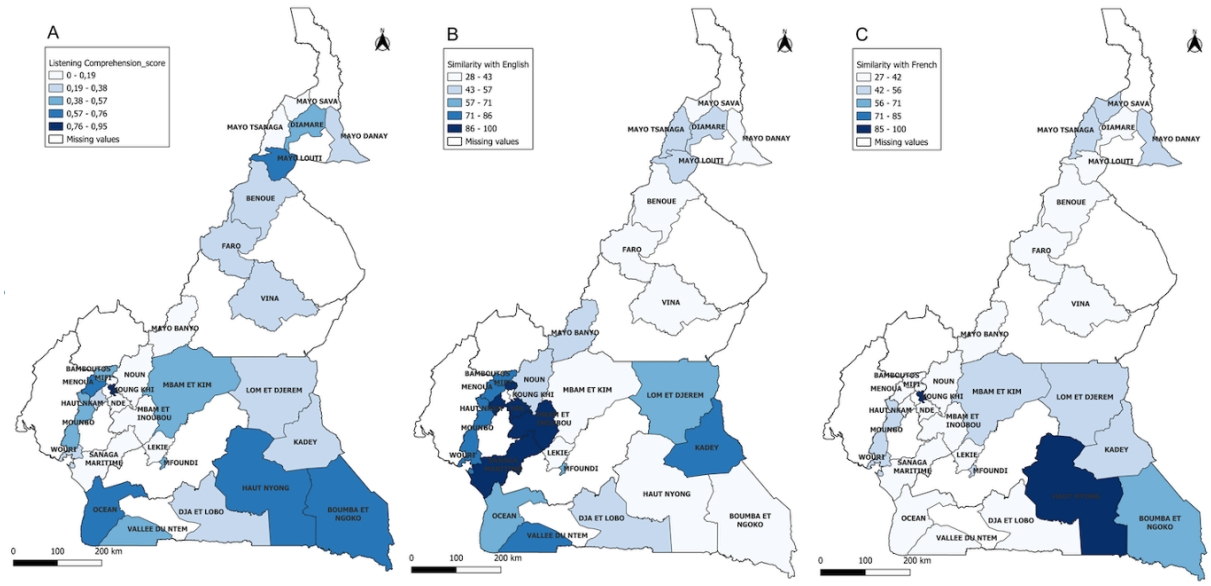
[2] The geographic distances (in kilometers) are summarized in Table A3 of the Appendix.

5.1 *Descriptive analysis*

Figure 1 visually represents spatial heterogeneity in both listening comprehension outcomes and linguistic similarity with the official languages across Cameroonian provinces. The figure contains three subfigures (maps of Cameroon), each corresponding to a distinct variable. Panel A shows the average listening comprehension test scores by province. Panel B displays the average phonetic similarity between the local languages and English, while Panel C does the same for French. These maps are constructed using 2019 EGRA test scores and lexicostatistical similarity indices. Darker colors indicate higher values: in Panel A, darker provinces reflect stronger average performance in listening comprehension, while in Panels B and C, they reflect greater phonetic similarity to the LOI.

Provinces such as Menoua, Diamaré, and Koung-Khi appear to record relatively higher average scores. In contrast, regions like Mayo-Louti, Vina, and parts of the East exhibit much lower scores, pointing to significant comprehension skills challenges. The right-hand map shows similarly uneven patterns of linguistic similarity. In several francophone regions, especially in the Centre, West, and South, the similarity index with French is relatively high, indicating stronger phonetic proximity between local languages and the official language of instruction. Conversely, several in the Far North and East record lower values.

Figure 1: Mean of Listening Comprehension Test Scores (A) and Linguistic Similarity with English(B) and French (C) by Province



Notes: **Source:** author. **Notes:** The maps are computed using 2019 EGRA data and linguistic data. Darker colors indicate stronger phonetic similarity or higher listening comprehension scores. Only provinces with available data are displayed. The maps use the Transverse Mercator projection with the standard WGS 1984 UTM Zone 33N coordinate system.

While a systematic positive correlation between linguistic similarity and listening comprehension scores is not immediately evident from the spatial maps, Figure 2 provides a clearer view

of this relationship through scatter plots. The figure is composed of two panels. The top panel (A) plots the average listening comprehension score per province against the mean similarity with French; the bottom panel (B) does the same for English. Each dot represents a province, and the blue line shows a linear fit. Dashed horizontal and vertical lines indicate national averages in listening scores and linguistic similarity, as reported in the summary statistics (Table 2).

Overall, a positive association is evident in both panels, although it seems more pronounced for French. In Panel A, provinces such as Koung-Khi, Haut-Nyong, and Boumba-et-Ngoko exhibit high similarity with French along with above-average listening comprehension scores. In the case of Koung-Khi, for instance, this may be partly explained by the local language context. One of the major indigenous languages spoken there is Nda'nda', which is second in prevalence after *Nghomala'*, and was not included in the similarity analysis due to data limitations. According to the similarity estimates presented in Table 1, Nda'nda' ranks among the languages that are most phonetically similar to French, with a similarity score of 50. This relatively high alignment may help explain why Koung-Khi records the highest average listening comprehension scores across all provinces.

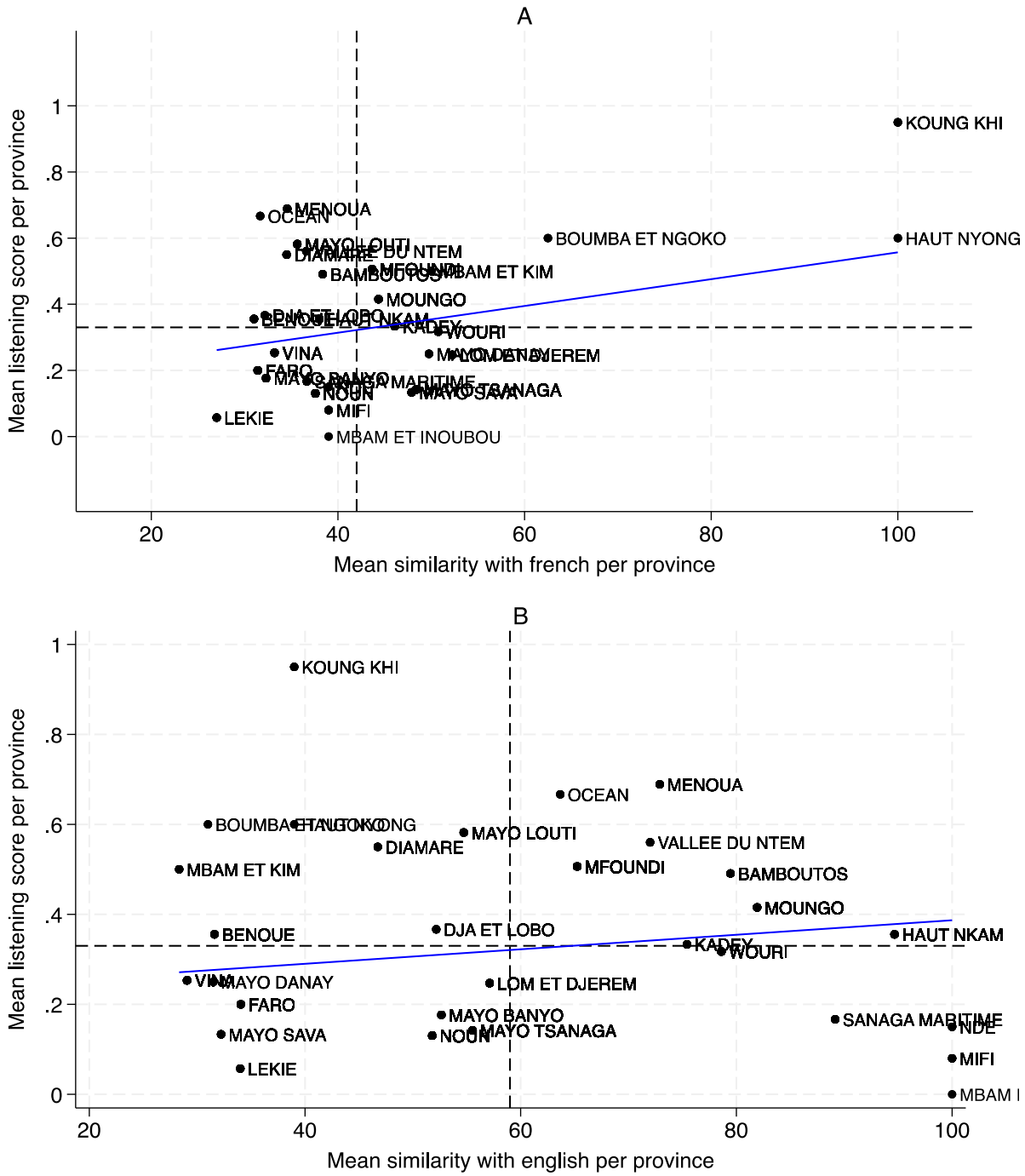
For English, the relationship seems less systematic. Many provinces show no consistent patterns. For instance, Mifi and Sanaga-Maritime exhibit relatively high similarities with English but only modest listening scores, suggesting that other contextual or educational factors may influence the relationship.

5.2 *Linguistic Similarity and Listening Comprehension: Main Estimation Results*

This section presents the main estimation results from the 2SLS regressions assessing the effect of linguistic similarity on listening comprehension outcomes. Table 3 is structured into four columns, reporting separate specifications for similarity with French and English. For each LOI, the first-stage regression reports the effect of geographic instruments (log distance to relevant cities) on phonetic similarity, while the second-stage regression reports the estimated impact of predicted similarity on listening comprehension scores. Control variables are included in all specifications and consist of student characteristics, household literacy and asset indicators, and province fixed effects.

Regarding the similarity with French, a one-point increase in linguistic similarity is associated with a 0.012 standard deviation increase in comprehension scores. A comparable and statistically significant effect is observed with the similarity to English, with a slightly smaller coefficient of 0.008. These improvements correspond to gains of about 3.7% and 2.5%, respectively, relative to average student performance. These findings confirm that the linguistic distance between L1 and the LOI constitutes a meaningful constraint on foundational learning.

Figure 2: Relationship Between Linguistic Similarity and Listening Comprehension Across Provinces



Notes: **Source:** Author. **Note:** The scatter plots show the average listening comprehension score by province plotted against the mean linguistic similarity with the language of instruction, French (panel A) and English (panel B). The blue line depicts a linear fit to the data. The dashed horizontal and vertical lines indicate the national averages of listening scores and linguistic similarity, respectively, as reported in the summary statistics of Table 2.

This result aligns with evidence from multilingual contexts suggesting that greater familiarity with the LOI facilitates the acquisition of oral comprehension skills in early schooling (Piper et al., 2016; Heugh et al., 2007). In Cameroon, most children enter school without prior exposure to French or English. In such contexts, phonetic proximity helps children process sounds more easily and build connections between spoken and written language, especially when decoding new vocabulary. This lowers the cognitive effort required to follow instructions, especially when literacy foundations are built in early grades. Although the average linguistic similarity with English is slightly higher than that with French (as shown in Table 2), the estimated return to phonetic similarity appears marginally greater for French. One possible explanation is that several widely spoken indigenous languages (e.g., Ewondo, Bulu, Dii) are structurally closer to French than to English, particularly in phonotactic and prosodic features. Additionally, variation in the sociolinguistic environments associated with the two LOIs may contribute to these differences. For instance, regions where French is the LOI often exhibit more stable language exposure patterns and a higher concentration of trained teachers, potentially reinforcing the impact of phonetic proximity.

The results also highlight the importance of household characteristics. Maternal literacy is positively and significantly associated with listening comprehension, reflecting well-established findings on the intergenerational transmission of language skills. Access to radios and televisions is positively associated with performance, suggesting that informal exposure to the LOI through media complements formal instruction. Conversely, age is negatively associated with comprehension outcomes, consistent with evidence from the region indicating that over-age enrollment, often a result of delayed school entry or grade repetition, can hinder learning in early grades.

Standard diagnostics support the identification strategy. The F-statistics for the first-stage regressions are 10.67 and 13.07, exceeding the conventional rule-of-thumb threshold of 10 proposed by Staiger and Stock (1997). The Hansen J-statistics for over-identification are not statistically significant (p-values of 0.585 and 0.430), which provides suggestive evidence that the instruments are not correlated with the second-stage error term.

5.3 *Heterogeneity Analysis by Education System, Locality, and Home Learning Environment*

Having established the effects of linguistic similarity, this section examines how these effects vary across educational subsystems, localities, and household learning contexts. Table 4 presents the results of the heterogeneity analysis with multiple panels. Specifically, the table disaggregates the sample along four dimensions: education system (Francophone vs. Anglophone), locality (urban vs. rural), home reading practices (read vs. don't read at home), and homework support mechanisms (help from parents, siblings, or no one). Each panel estimates the effect of

Table 3: Regression Results of the Effect of Linguistic Similarity on Listening Comprehension Skills

	Outcome variable = Listening comprehension score			
	First stage	Second stage	First stage	Second stage
Similarity with French		0.012**		
		(0.006)		
Similarity with English				0.008**
				(0.004)
Age	−0.105	−0.037**	−0.231**	−0.019
	(0.992)	(0.018)	(0.116)	(0.020)
Sex (Female)	0.101	−0.002	−0.613**	0.051
	(0.239)	(0.045)	(0.302)	(0.050)
Father is literate	0.002	0.020	0.305	0.002
	(0.007)	(0.057)	(0.394)	(0.057)
Mother is literate	0.273***	0.146**	0.365	0.139**
	(0.067)	(0.063)	(0.429)	(0.065)
Radio at home	0.092	0.119**	0.106	0.114**
	(0.112)	(0.047)	(0.305)	(0.046)
Electricity at home	0.268*	0.104*	0.279	0.097*
	(0.162)	(0.053)	(0.345)	(0.053)
TV at home	0.226	0.100*	−0.245	0.129**
	(0.205)	(0.058)	(0.360)	(0.054)
Computer at home	−0.099	−0.100	0.245	−0.087
	(0.123)	(0.063)	(0.407)	(0.065)
Reads at home	0.050*	0.170***	0.369*	0.176***
	(0.030)	(0.065)	(0.223)	(0.067)
Parents help with homework	0.093***	0.051*	0.035	0.034**
	(0.021)	(0.119)	(0.071)	(0.017)
Education subsystem (Francophone)	0.259***	0.040**	0.088**	0.024
	(0.065)	(0.019)	(0.044)	(0.033)
Locality (Rural)	−0.088**	−0.060***	0.052	−0.095***
	(0.044)	(0.019)	(0.101)	(0.016)
Log distance to Yaoundé	0.070**			
	(0.035)			
Log distance to Douala	0.088***			
	(0.024)			
Log distance to Bamenda			0.042***	
			(0.014)	
Log distance to Buea			−0.040***	
			(0.011)	
N		449		449
R^2		0.558		0.661
F-stat		10.67		13.07
P-value (Hansen)		0.585		0.430

Notes: Source: Author's calculations. All models include fixed effects for the province where the school is located. The instruments excluded from the 2SLS regressions correspond to the distances between students' place of residence and the cities with the highest proportion of French or English speakers. The instruments for similarity with French are the distances to Yaoundé and Douala, while the instruments for similarity with English are the distances to Bamenda and Buea. The F-test assesses the relevance of the instruments, and the Hansen p-value tests overidentification. Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

phonetic similarity on listening comprehension scores while controlling for individual, household, and school-level characteristics.

The results suggest that linguistic similarity significantly enhances listening comprehension, particularly in Francophone schools. In Anglophone schools, the effect is weaker but still marginally significant for students whose mother tongue is phonetically closer to English. This discrepancy could be attributed to fundamental differences in language policies between the two educational systems. The Francophone system follows a strict monolingual approach, where students are introduced to French as the exclusive LOI from an early stage and expected to master it without transitional bilingual support. As a result, students whose mother tongue is phonetically close to French may experience a natural advantage in comprehension, as they can more easily map their existing phonetic structures onto the LOI. By contrast, the Anglophone education system is characterized by greater linguistic flexibility, with more tolerance for code-switching and exposure to multiple languages in school and daily life (Bamgbose, 2011). This could dilute the direct impact of linguistic similarity, as students develop a broader multilingual repertoire rather than relying solely on LOI proximity. However, it is important to acknowledge that the two predominantly Anglophone regions of Cameroon (Northwest and Southwest) were excluded from the EGRA assessment due to ongoing instability. While this limits direct comparisons, it does not entirely undermine the findings, as Anglophone schools exist in the other eight regions, albeit in smaller numbers compared to Francophone schools. Recent findings by Bekkouche and Dupraz (2023) reinforce this perspective, highlighting an unexpected advantage of the Francophone system in primary education outcomes. Although the authors did not pinpoint specific mechanisms driving this trend, the current study suggests that linguistic similarity may be an explanatory factor.

Striking differences also emerge by locality. In rural areas, phonetic similarity with the LOI has a markedly stronger effect on listening comprehension outcomes than in urban areas. While the coefficients remain statistically significant in both settings, the magnitude of the impact is substantially higher in rural communities. This pattern suggests that linguistic proximity plays a more compensatory role in rural environments, where children are less likely to be exposed to the LOI through media, community interactions, or extracurricular resources. In urban areas, by contrast, students may benefit from greater incidental exposure to the LOI, through television, radio, signage, or interactions in formal and informal settings, which can partially offset the disadvantage of phonetic dissimilarity. As a result, the marginal benefit of linguistic similarity is lower in these environments. In rural settings, however, where exposure to the LOI is often limited to the classroom, children whose L1 is phonologically close to the LOI are better positioned to grasp oral instructions, retain vocabulary, and build early comprehension skills.

The analysis of home reading practices further supports the idea that linguistic similarity alone is not enough it must be reinforced by regular exposure through reading. Students who

read at home experience greater benefits from linguistic similarity with French, while those who do not engage in home reading still show a positive but weaker effect. This highlights the complementary role of early literacy exposure in reinforcing phonetic advantages. Empirical studies confirm this pattern: Pretorius and Mampuru (2007) found that early exposure to reading at home enhances phonological awareness, which in turn improves literacy skills in school. Similarly, Piper et al. (2018) in Kenya demonstrated that students with access to books at home, particularly in languages closely related to the LOI, exhibited stronger reading comprehension skills.

Interestingly, linguistic similarity with English has a marginally significant effect among students who read at home, but for those who do not engage in reading, the effect is either insignificant or negative. This finding aligns with research by Laitin et al. (2019), which shows that merely being exposed to a linguistically distant LOI is not sufficient active interaction through reading is necessary to bridge phonological gaps. English, being phonetically and structurally more distant from most Cameroonian indigenous languages than French, may require more intensive reinforcement through home reading for linguistic similarity to translate into tangible literacy benefits. These results suggest that linguistic proximity alone does not guarantee improved literacy outcomes unless supported by consistent reading exposure.

Finally, the analysis of homework support highlights the critical role of parental and sibling involvement in amplifying the effects of linguistic similarity. Students who receive homework assistance from parents exhibit strong positive effects, while those who lack support fail to benefit from linguistic similarity. This aligns with the findings of Desforges and Abouchaar (2003), who emphasize that parental involvement is one of the strongest predictors of academic success, particularly in language acquisition. The effect is especially pronounced when linguistic similarity with French is combined with parental support, suggesting that structured family assistance enhances the cognitive processing of phonetic similarities between the home language and the LOI. For linguistic similarity with English, parental help has a positive and significant effect, while sibling assistance appears weaker and statistically insignificant. This discrepancy suggests that parental involvement provides structured reinforcement, whereas peer or sibling support may lack consistency or pedagogical depth. Research by Evans and Shaw (2008) in multilingual environments confirms that parental involvement in language-related homework tasks significantly improves children's phonological awareness and comprehension skills, reinforcing the idea that active parental engagement amplifies the benefits of linguistic similarity.

5.4 *Can Linguistic Similarity Offset Regional Disparities?*

Finally, this section investigates whether linguistic similarity with the LOI reduces regional disparities in listening comprehension. As shown in Figures 1 and 2, there is substantial geographic variation in both phonetic similarity and listening outcomes. To assess whether greater align-

Table 4: Heterogeneity Analysis by Education Subsystem, Locality, Family Learning Environment, and Homework Support Mechanisms

	Outcome variable = Listening comprehension score					
Panel A	Anglophone subsystem		Francophone subsystem			
	(1)	(2)	(3)	(4)		
Similarity with French	0.078		0.089**			
	(0.099)		(0.045)			
Similarity with English		0.012*		0.189**		
		(0.007)		(0.099)		
Controls	Yes	Yes	Yes	Yes		
R^2	0.133	0.322	0.355	0.221		
N	127	127	263	263		
P-value (Hansen)	0.322	0.167	0.121	0.267		
Panel B	Urban		Rural			
Similarity with French	0.008**		0.018**			
	(0.004)		(0.009)			
Similarity with English		0.005*		0.015**		
		(0.003)		(0.007)		
Controls	Yes	Yes	Yes	Yes		
R^2	0.322	0.406	0.211	0.177		
N	226	226	165	165		
P-value (Hansen)	0.205	0.089	0.267	0.156		
Panel C	Read at home		Don't read at home			
Similarity with French	0.039**		0.026*			
	(0.015)		(0.016)			
Similarity with English		0.008*		−0.029		
		(0.004)		(0.022)		
Controls	Yes	Yes	Yes	Yes		
R^2	0.132	0.052	0.505	0.402		
N	246	246	124	124		
P-value (Hansen)	0.145	0.097	0.204	0.056		
Panel D	Help with homework					
	Parents	Siblings	No one	Parents	Siblings	No one
Similarity with French	0.046**	0.107	−0.134			
	(0.018)	(0.106)	(0.150)			
Similarity with English				0.028*	−0.003	0.007
				(0.016)	(0.029)	(0.012)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.620	0.100	0.058	0.369	0.232	0.152
N	84	62	62	88	64	62
P-value (Hansen)	0.166	0.451	0.052	0.192	0.031	0.081

Notes: **Source**: author. **Note**: The second stage of the IV-2SLS estimates are reported. The instruments for similarity with French are the distances to Yaoundé and Douala, while the instruments for similarity with English are the distances to Bamenda and Buea. Control variables are included but not displayed. Standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

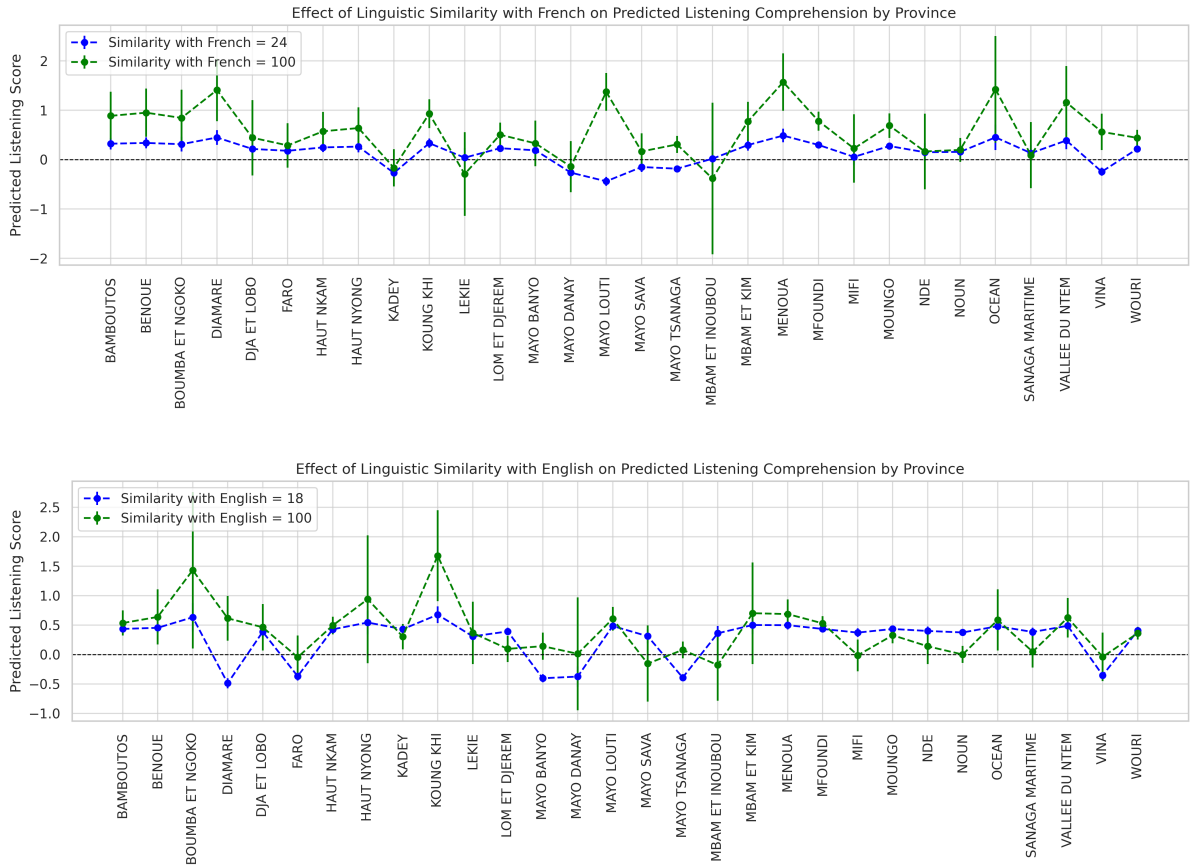
ment between students' L1 and the LOI lessens this spatial inequality, I estimate predictive margins of listening comprehension across provinces under contrasting levels of linguistic similarity.

Specifically, I use the minimum and maximum values of linguistic similarity as reported in Table 2 to define the two scenarios. For French, the minimum observed value is 24 and the maximum is 100. For English, the minimum is 18 and the maximum is also 100. These extreme values represent the lowest and highest levels of phonetic proximity recorded in the sample. The margins are computed using the second-stage 2SLS specification described in Section 5, with listening comprehension as the dependent variable and linguistic similarity (with either French or English) as the instrumented regressor. All control variables from the main analysis are included. Figure 5 presents the results. The top panel displays predicted listening comprehension scores across provinces under low (minimum) and high (maximum) phonetic similarity with French. The bottom panel shows the same comparison for English. In each panel, provinces are represented by dots indicating predicted scores under each similarity condition, with dashed lines linking the low-similarity (blue) and high-similarity (green) cases. Confidence intervals at 95% are shown as vertical lines.

The results indicate that in most provinces, including Menoua, Diamaré, Koung-Khi, and Ocean, predicted comprehension scores improve considerably under high similarity to French. In several disadvantaged regions, such as Mayo-Louti and Vina, the predicted score under low similarity is negative but rises above zero when similarity is high. This suggests that linguistic similarity to French compensates for the improvement of foundational skills in underperforming areas.

The pattern is more variable for English. Gains in predicted scores under high similarity are significant only in a few provinces, such as Boumba-et-Ngoko, Koung-Khi, and Menoua. In many others, scores remain stable, or the confidence intervals overlap, indicating no statistically significant change. In isolated cases like Mayo-Banyo and Noun, the predicted score under high similarity slightly declines; however, the change is negligible and not statistically significant.

Figure 3: Effect of Linguistic Similarity with French and English on Predicted Listening Comprehension Across Provinces



Notes: Source: Author's calculations. **Note:** The graphs display predicted listening comprehension scores by province under two levels of linguistic similarity with the LOI (French or English), based on the minimum and maximum values observed in Table 2. A similarity value of 100 implies perfect phonetic alignment. Margins are computed using the second-stage IV model described in Section 5, with full covariate adjustment. Confidence intervals are shown at the 95% level.

6 Conclusion

This study examined the effect of linguistic similarity between students' home language and the language of instruction on listening comprehension skill in Cameroon. The results show that phonetic proximity to the LOI significantly improves comprehension, particularly in Francophone schools. However, the effect depends on context: it is stronger in rural areas and among students who benefit from reading support at home. In contrast, in urban settings or in households without reading practices, the benefits are weaker or absent.

These findings have important implications for education policy in Cameroon and other multilingual contexts. The first key takeaway is that linguistic distance matters, and other interventions aimed at improving access to early childhood education may not consistently achieve the expected results if they fail to address the language barrier many children face, particularly in rural areas. Educational policies should align the LOI more closely with students' linguis-

tic backgrounds, particularly in early-grade instruction. While full-scale mother tongue-based education programs may face logistical and political challenges, incremental reforms such as bilingual transitional programs, phonetic adaptation of textbooks, or targeted language support interventions could enhance learning outcomes. Additionally, promoting home literacy environments is crucial. Given the strong interaction between linguistic similarity and home reading practices, policies that expand access to books, promote parental literacy, and encourage reading habits at home could amplify the benefits of linguistic proximity. Expanding educational media programs, particularly in regions with low literacy rates, could be a low-cost intervention to strengthen language exposure outside school. Finally, regional disparities in education outcomes indicate that linguistic adaptation policies should be complemented by broader investments in teacher training, classroom resources, and community-driven language programs to ensure that students in marginalized areas receive adequate support.

This study's limitations include concerns about measurement errors in linguistic similarity. The absence of natural experiments in Indigenous language-in-education policies at the national level limits the ability to improve identification. Additionally, data coverage limitations prevent a direct analysis of Anglophone-majority regions (Northwest and Southwest), which were excluded from the EGRA dataset due to instability. Although the presence of Anglophone schools in the remaining eight regions allows for comparisons, a more comprehensive dataset covering all regions would provide a fuller picture of how linguistic similarity operates across both educational systems.

Nonetheless, this study opens several avenues for future research. First, more precise linguistic similarity measures, potentially incorporating phonological processing and cognitive load assessments, could provide a more nuanced understanding of how linguistic proximity influences literacy development. Second, longitudinal studies tracking students over time could help determine whether the benefits of linguistic similarity persist beyond early literacy development and extend to other domains such as writing, reading comprehension, and subject-specific learning (e.g., mathematics, science). Finally, given that linguistic similarity appears to mitigate learning disadvantages in underprivileged regions, future research could explore whether language-based interventions can serve as cost-effective tools for improving educational equity in multilingual developing countries.

References

Avom, D., Kuété, F. Y. M., & Kouamo, R. N. (2021). Does educational decentralization improve the technical efficiency of elementary schools in sub-Saharan Africa?. *Mondes en developpement*, *195*(3), 53-74.

Ball, J. (2010). Educational equity for children from diverse language backgrounds: Mother tongue-based bilingual or multilingual education in the early years. UNESCO.

Bamgbose, A. (2011). African languages today: The challenge of and prospects for empowerment under globalization. In *Selected proceedings of the 40th annual conference on African linguistics* (pp. 1-14). Somerville: Cascadilla Proceedings Project.

Bekkouche, Y., & Dupraz, Y. (2023). Colonial origins and quality of education evidence from Cameroon. *World Development*, 170, 106245. `https://doi.org/10.1016/j.worlddev.2023.106245`

Benson, C. (2004). The importance of mother tongue-based schooling for educational quality. EFA Global Monitoring Report.

Benson, C. (2005). The importance of mother tongue-based schooling for educational quality. EFA Global Monitoring Report 2005: The Quality Imperative.

Booth, J. R., Lu, D., Burman, D. D., Chou, T. L., Jin, Z., Peng, D. L., ... & Liu, L. (2006). Specialization of phonological and semantic processing in Chinese word reading. *Brain research*, *1071*(1), 197-207.

Bosker, M., & Buringh, E. (2017). City seeds: Geography and the origins of the European city system. *Journal of Urban Economics*, *98*, 139-157.

Chiatoh, B. A. (2012). Language education and policy in Cameroon: A critique of the implementation of the mother tongue in the educational system. Journal of Educational and Social Research, 2(2), 1–8.

Chimombo, J. P. G. (2009). Changing patterns of language of instruction in Malawi: Policy and practice. *Comparative and International Education*, 38(2), 37–48.

Chiswick, B. R., & Miller, P. W. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of multilingual and multicultural development*, *26*(1), 1-11.

Chiswick, B. R., & Miller, P. W. (2008). A test of the critical period hypothesis for language learning. *Journal of multilingual and multicultural development*, *29*(1), 16-29.

Cummins, J. (2000). Language, power, and pedagogy: Bilingual children in the crossfire. Multilingual Matters.

De Gregorio, J., Toral, R., & Sánchez, D. (2024). Exploring language relations through syntactic distances and geographic proximity. *EPJ Data Science*, *13*(1), 61.

Dekker, D., & Young, C. (2005). Bridging the gap: The role of local languages in education. International Journal of Bilingual Education and Bilingualism, 8(4), 339-354.

Desforges, C., & Abouchaar, A. (2003). *The impact of parental involvement, parental support and family education on pupil achievement and adjustment: A literature review* (Vol. 433). London: DfES.

Dieu M. and P. Renaud, dir. (1983), Atlas linguistique du Cameroun ALCAM : Inventaire préliminaire, Paris / Yaoundé : ACCT / CERDOTOLA. https://comparalex.org/index.php?page=stdlist&id=3

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2024). Ethnologue: Languages of the World (27th ed.). SIL International.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2023. Ethnologue: Languages of the World. Twenty-sixth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Evans, M. A., & Shaw, D. (2008). Home grown for reading: Parental contributions to young children's emergent literacy and word recognition. *Canadian Psychology/Psychologie Canadienne*, *49*(2), 89.

Fafunwa, A. B., et al. (1989). Education in mother tongue: The Ife primary education research project. University Press Limited.

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of economic growth*, *8*, 195-222.

Glewwe, P., & Muralidharan, K. (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In *Handbook of the Economics of Education* (Vol. 5, pp. 653-743). Elsevier.

Heugh, K., Benson, C., Bogale, B., & Yohannes, M. A. (2007). Study on Medium of Instruction in Primary Schools in Ethiopia. Ministry of Education, Ethiopia.

Igarashi, T., Maulana, S., & Suryadarma, D. (2024). Mother tongue-based education in a diverse society and the acquisition of foundational skills: Evidence from the Philippines. *Labour Economics*, *91*, 102641.

Kim, Y. S., & Phillips, B. (2014). Cognitive correlates of listening comprehension. *Reading Research Quarterly*, *49*(3), 269-281.

Kouamo, R. N., & Dieng, A. (2022). The determinants of the quality of primary education in Cameroon. *The Journal of Quality in Education*, *12*(20), 21-36.

Laitin, D. D., Ramachandran, R., & Walter, S. L. (2019). The legacy of colonial languages: Unequal access to education and literacy outcomes. Economic Development and Cultural Change, 68(1), 211-247.

List, J. M. (2012). LexStat: Automatic detection of cognates in multilingual wordlists. In *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH* (pp. 117-125).

List, M. (2014). *Sequence comparison in historical linguistics* (Vol. 1). Walter de Gruyter GmbH & Co KG.

McMahon, A., & McMahon, R. (2005). *Language classification by numbers*. Oxford University Press.

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. *American Economic Review*, *102*(4), 1508-1539.

Nakagawa, M., & Sugasawa, S. (2022). Linguistic distance and economic development: A cross-country analysis. *Review of Development Economics*, *26*(2), 793-834.

Nana, G. (2013). This is No French School: Language and Education Traditions in Primary Schooling in Cameroon—A Comparative Perspective. *Research in Comparative and International Education*, *8*(4), 468-509.

PASEC (Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN) (2019). PASEC 2019: Education System Performance in Francophone Africa. Dakar, Senegal: CONFEMEN.

Perfetti, C.A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven. C. Elbro, & P. Reitsma (Eds.), Precursors of Functional Literacy Amsterdam/Philadelphia: John Benjamins.

Piper, B., Zuilkowski, S. S., & Ong'ele, S. (2016). Implementing mother tongue instruction in the real world: Results from a randomized trial in Kenya. Comparative Education Review, 60(4), 776–807.

Pretorius, E. J., & Mampuru, D. M. (2007). Playing football without a ball: language, reading and academic performance in a high-poverty school. *Journal of Research in Reading*, *30*(1), 38-58.

Schepens, J., Van Hout, R., & Jaeger, T. F. (2020). Big data suggest strong constraints of linguistic similarity on adult language learning. *Cognition*, *194*, 104056.

Serva, M., & Petroni, F. (2008). Indo-European languages tree by Levenshtein distance. *Europhysics letters*, *81*(6), 68005.

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *science*, *328*(5977), 450-452.

Staiger, D., & Stock, J. (1997). to Econometrica. *Econometrica*, *65*(3), 557-586.

Tadadjeu, M. (2003). Bilinguisme identitaire et apprentissage d'une troisième langue: le cas du Cameroun. *African Journal of Applied Linguistics*, *4*, 5-12.

Tadadjeu, M. (2004). Language, literacy and education in African development: A perspective from Cameroon. *SIL International*, *1*(10).

Tadadjeu, M. & Sadembouo, É. (2000). PROPELCA Reports. Yaoundé: CNEAL.

Trudell, B. (2005). Language choice, education and community identity. *International Journal of Educational Development*, *25*(3), 237-251.

Trudell, B. (2009). Local-language literacy and sustainable development in Africa. International Journal of Educational Development, 29(1), 73–79.

UNESCO (2021). Understanding learning and its trajectory in multilingual contexts. Paris: UNESCO.

UNESCO (United Nations Educational, Scientific and Cultural Organization) (2022). International Decade of Indigenous Languages 2022–2032: A Global Action Plan. Paris:

Walter, S., & Benson, C. (2012). Language policy and medium of instruction in Africa: The key to sustainable development. Comparative Education, 48(1), 75-90.

Weiss, T. L. (1996). *Les migrations nigérianes dans le sud-ouest du Cameroun* (Doctoral dissertation, Paris 4).

Wichmann, S., Müller, A., & Velupillai, V. (2010). Homelands of the world's language families: A quantitative approach. *Diachronica*, *27*(2), 247-276.

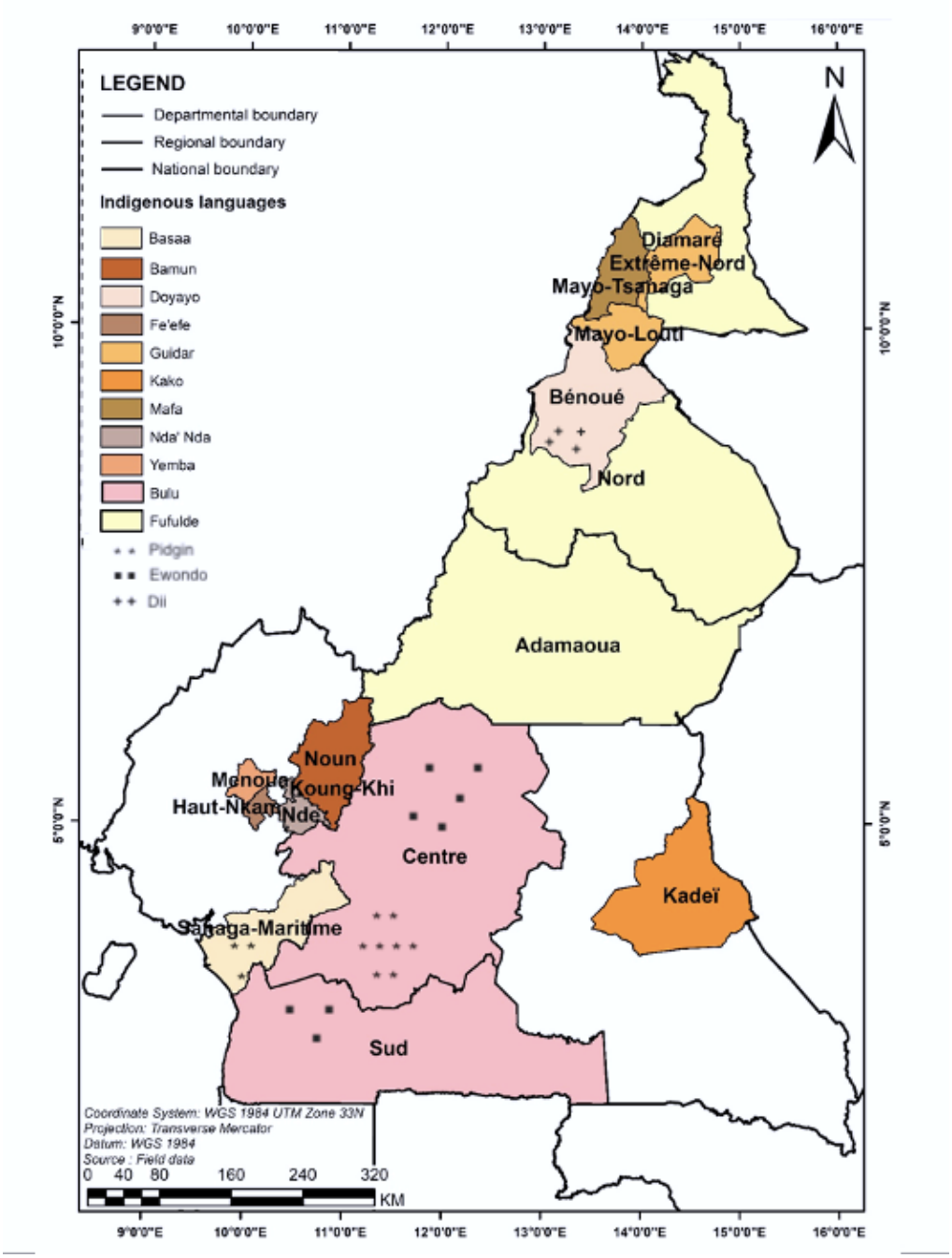
Williams, E., and Ball, J. (2014). *Language and Literacy in Rwanda: Learning to Teach in the Mother Tongue*. British Council.

World Bank (2020). World Development Indicators: Cameroon education profile.

World Bank. (2018). Learning to Realize Education's Promise: World Development Report 2018. Washington, DC: World Bank.

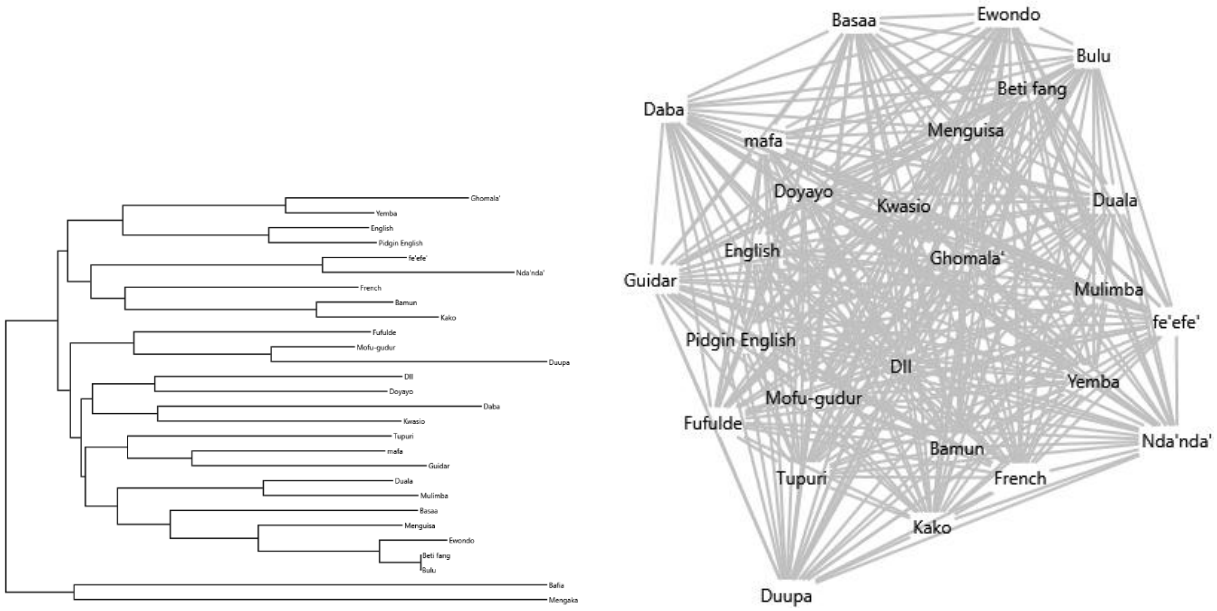
Appendix A: Additional tables and figures

Figure A1: Distribution of retained indigenous languages across country



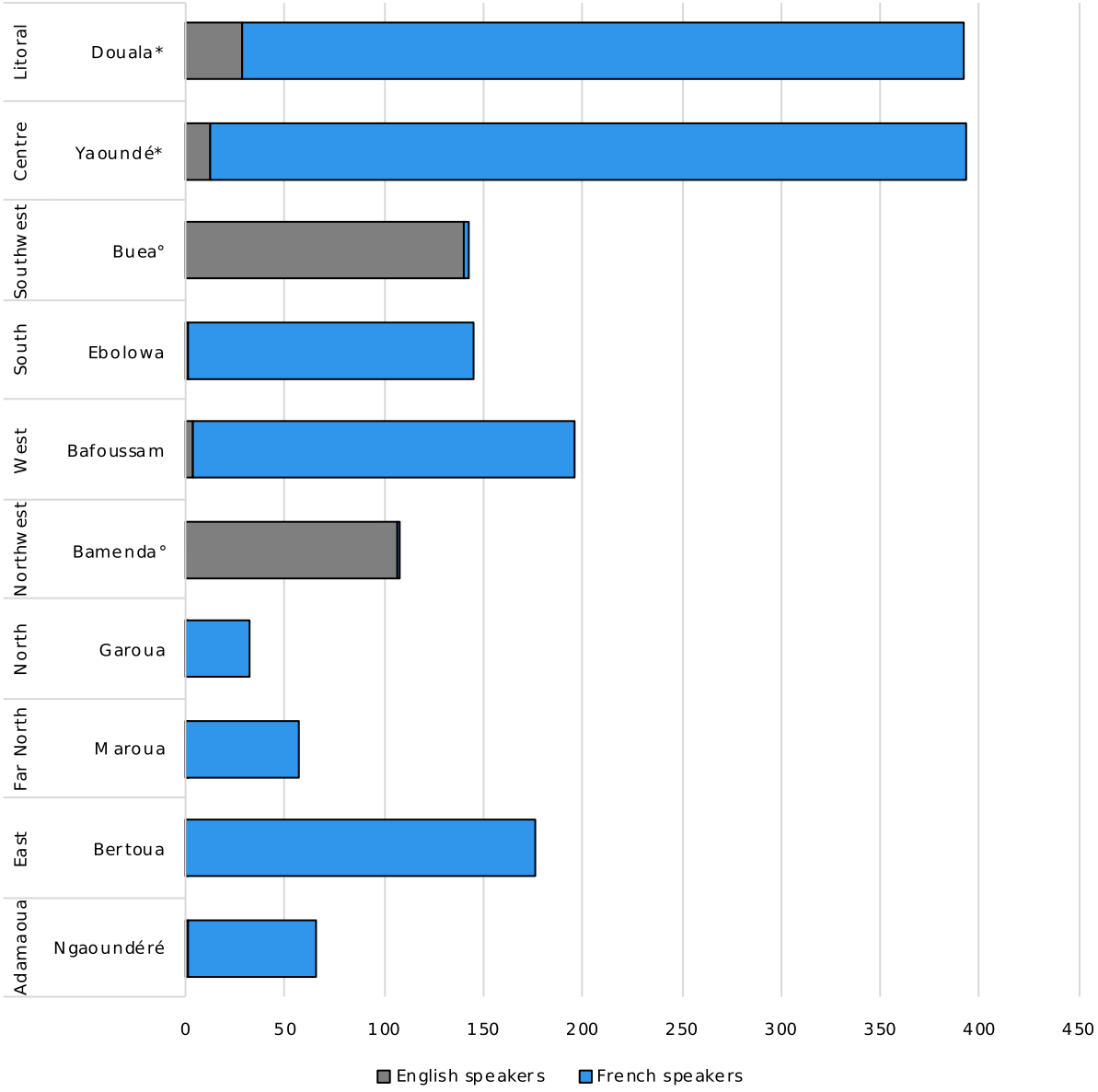
Notes: **Source:** Author, based on field data and Ethnologue (2010). The map uses the Transverse Mercator projection with the World Geodetic System 1984 (WGS84) UTM Zone 33W coordinate system. **Note:** The distribution reflects the proportion of speakers of each language by province and region, as reported in Ethnologue. Some languages, such as Ewondo, Bulu, and Dii, are spoken across multiple provinces. Pidgin, predominantly spoken in the Northwest and Southwest regions (excluded from this study), is also present in the Centre and Littoral regions due to internal migration.

Figure A2: Phonetic dendrogram and network between languages



Notes: **Source**: Author. **Notes**: The phonetic dendrogram (left) represents the hierarchical clustering of languages based on their phonetic similarity, illustrating how closely related different indigenous languages are to each other and to the languages of instruction (French and English). The phonetic network (right) visualizes the pairwise phonetic distances between languages, with stronger connections indicating higher similarity. These representations are based on lexicostatistical analysis using data from the 2019 EGRA survey and processed with COG software.

Figure A3: Distribution of LOI Speakers Across Regional Capital Cities



Notes: **Source**: Author's calculations based on the 2018 Cameroon Demographic and Health Survey (DHS).

Table A1: Lists of words used to calculate linguistic similarities and their translations

List ID	List Gloss English	List Gloss French	Bamun	Pidgin English	Bulu	Ewondo	Fufulde	Yemba	Daba
1	mouth	bouche	?u?t	mop	anyu	anyu	Houndouko	nchu	M?
2	eye	oeil	li	ay	Di?s	Di?s	Gwté	lez??k	h??rya
3	head	tête	tu	head	n?lo?	n?lo?	Hore	athu?	Tálán
4	hair	cheveux (de tête)	?utu	biebie	Bisi?l (ya? nlo?)	Bisi?l ya? nlo?	Gasa Hore	nts?'	Sòson tálán
5	tooth	dent	zu	tid	Aso??	Aso??	Yihe	les??	nd?n
6	tongue	langue	lum	tong	Oy??m	Oy??m	Dengal	al??	gànà
7	nose	nez	ji?	nos	Dzo?é	Dzwé	Kiné	lezo?ho?	mìc?n
8	ear	oreille	tut	iey	Al??	Al??	Noppi	letuu??	zlìm?
9	neck	cou	lut	nek	ci??	Ki??	Dade	nt???	w??là
10	breast	sein, mamelle	bun	bobi	Abé, énya??	Ab??	Hendi	lep??	Wà
11	hand, arm	main	buo	han	W??	W??	Djoungo	apu?	k??rà
12	hand, arm	bras	buo	han	Enam	Akul	Djokké	lekw??	k??rà pay
13	claw, fingernail	griffe	kew??	finga	Byi?a?y	Byé	Gasa	??kuu?be	k??r pìz?n
14	claw, fingernail	ongle	kew??	han nail	Di?a?y	Dzi?é	Pedeli	?kuu?p	Pìz?n
15	foot, leg	pied	kut	foot	Ab?	Ako?l	Kosde	akhu	Ngàz
16	foot, leg	jambe (y compris le pied)	kut	foot	Abe (ab?)	Ab? (ako?l)	Kosde	akhu	Ngàz p?y
17	foot, leg	jambe	kut	foot	Abe	Ab?	Korle	nta?	Ngàz p?y
18	buttock	fesse	tam?it	butoks	Aka?n	Aka?n	Pobe ou Boundou	mesakats??	Màdàb?
19	belly	ventre	vum	bele	Abum	Abum	Redou	lev?m	Lìb?
20	belly	ventre, abdomen (externe)	vum	bele	Abum,ongon	Abum(ongon)	Beche Derredou	ndo?lev?m	Màdàb? lìb?
21	navel	Nombril	tu?	nevel	D??p	D??b	Djabourou	let??	Bàbràv
22	intestines	intestins	two	intestans	Min?ya	Min?y?	Téteki	metu	Wej?
23	blood	Sang	?i	blood	M?tii?	M?ki?	iam	metshii?	?è?iz
24	urine	Urine	?e?ke	pis	M?nyo?lo?k	M?ny??l??g	Sillé	ntsents???	??z
25	bone	Os	ku?	bone	Ev??s	Eve?s	Iyal	akw??	k??rklin
26	skin	peau (d'homme)	kup	skin	Ekop (mot)	Ekob	Laral gorko	?gup-nn??t	Bòkùlà
27	wing	Aile	ba???	feda	Asa?	Asaa?	Biledji	lepap	Kòpòk
28	feather	Plume	Vy'	feda	Afa?p	Afa?b	Biledji	lepap	Kpètè
29	horn	Corne	duu?	horn	Min?lak	Min?lag	Loual	nd??	Fàl?m
30	tail	Queue	ku?n	tey	Ndj??m	Ng??m	Bicho	asa?	k?t?r
31	person / human	personne	m?n	man	Mot	Mod	Goddo	nyi?	Hìd?

(continued)

List ID	List Gloss English	List Gloss French	Bamun	Pidgin English	Bulu	Ewondo	Fufulde	Yemba	Daba
32	man (male)	homme (mâle)	m?ba	man pikin	Mot (n?no?m)	Mod (n?no?m)	Gokko	nyi?-mba?a	m?w?l
33	woman	Femme	m?bwi?	woman	Min?ga?	mininga?	Debbo	menzwi?	Wertì
34	husband	mari, époux	muri	massa	N?no?m, n?no?m	n?no?m	Gorkoma	ndo?	m?w?l
35	child (offspring)	Enfant	m?n	pikin	M??n	M??n	Binguell	m??	k?rtì
36	name	Nom	li?	name	Eyole	Dzo?é	Inde	lez???	Zlìm?
37	sky	Ciel	pu?t	sky	Yo?p	Dzo?b	Doulé	lepuhu	w?zlàv
38	night	Nuit	ba???	night	Alu?	Alu?	Djemma	es????-njh??	Vùd?
39	moon	Lune	mu?t	moon	Ng?n	Ng?n	Lewron	sa?	t?rà
40	sun	Soleil	?u?m	sun	Vya?n	Vya?n	Nague	nnu	Pìc
41	wind	Vent	fufwet	win	M?f?b?	M?f?b?	Hendou	ef?k	b??
42	cloud	Nuage	Ba?a'	clowd	N?ku?t	N?ku?d	Souddi	al?	Gàzlàv?y
43	dew	Rosée	zyn??	dew	M?fo	M?fo	Bichadjan	metswhi?	Mìnmim?
44	rain	Pluie	bu	rain	Mv??	Mv??	Yedé	mb??	V?n
45	earth (soil, ground)	sol, terre	sie?	grong	Si?, tia	Si?, m?ts??l??g	Leddi	esi?	Màyìk, tálán Màyìk
46	earth (soil, ground)	terre (sol)	sie?	grong	Tia (tia)	M?ts??l??g	Leddi	esi?	tálán Màyìk
47	sand	Sable	suen?a	sansan	N?s??l?k	N?s??l?g	Djarendi	asa?a?sa?a?	v??ga?
48	path	chemin, route	me?	road	Z??n, n?dj???	Ze?n	Lawol	n?z?h??	m?l?kotuf
49	water	Eau	kie	wata	M?ndi?m	M?ndi?b	Ndiam	ntshii?	y??m
50	stream, river	rivière, fleuve	kie	wata and bik wata	N?lot swé, oswé	N?lod oswé, oswé	Mayo, Mayo-Tekou	ndu	Màzà
51	stream, river	Ruisseau	kie		Ototo?	Ototo?	Mayel	m?nsthi?t	Màzà met??z
52	house	Maison	ndap	has	Nda?	Nda?	sare	?gya?	Jìk
53	house	concession, maison	?y	has	Nda?	Nda?	Nyibouso sare	mba	Dee
54	fire	Feu	mu?	fire	Ndu?an	Ndu?an	Yite	mmo?k	Kùh?
55	firewood	bois à brûler	kpi	firewud	Ndjaa ba? di?	Ndzaag ba? di?gi	Ledde woulougo	ntswi?	c?c?lem
56	smoke	Fumée	?i ?i	smoke	Oti?taa	Oti?daa	Sourde	??gihi?	Sèlèh
57	ash(es)	Cendres	vy	achis fire	Asu?p	Asu?b	Dodde	avoho?	Hìlèc?n
58	knife	Couteau	Me?i	nife	Oc??	Ok??	Labi	pu'	Wàz
59	rope	Corde		rop	N?k??l	N?k??l	Boggol	?kh??	mb?sèh
60	spear	lance (javelot)	ku?t	spear	Ak???	Ak???	Djaral	ama'	Njervèn
61	war	Guerre	ket	war	Bita?	Bita?	kono	ncu	j?g?m
62	meat, animal	Animal	pit	animal	ci?t	Tsi?d	Ndabbawa	na?	t?k??n gaz maa fo?
63	meat, animal	Viande	me?am	mit	Tsi?t	Tsi?d	Koussel	mbap	v?r?n
64	dog	Chien		dog	Mvu?	Mvu?	Goyrou	mbh??	Hazà
65	elephant	éléphant	?ap	elefan	Z?k	Z?g	Yiwa	esh??	bi?gìni

(continued)

List ID	List Gloss English	List Gloss French	Bamun	Pidgin English	Bulu	Ewondo	Fufulde	Yemba	Daba
66	goat	Chèvre	m?f?	goat	Ka?bat	Ka?bad	Bedva	mvho?	g?nàw
67	bird	Oiseau	zuen	bird	On??n	On??n	Solli	s???	V?kin
68	tortoise (land)	tortue (terrestre)	m?fi	totis	Ku?lu?	Ku?lu		mpfoa?k??'	Tàkk?l
69	snake	Serpent	m?si	snake	Ny??	Ny??	Bodi	nnu	v?n
70	fish	Poisson	m?tito'	fish	K??s	K??s	Liddi	esoho?	Kìl?f
71	louse	pou (de tête)	?uo		Nyi?n (ya? n?lo?)	Nyi?n (ya? n?lo?)	Tendi ùhore	f???	m??rcit
72	egg	Œuf		egg	Atyii?	Aki?	Guerade	lepo?gap	d?l?i
73	tree	Arbre	?a	tree	Elé	Elé	Leggal	ath?	Pay
74	bark (tree)	écorce (d'arbre)	gwotn?	canda stik	Eba?b (élé)	Eba?b élé	Cebre Leggal	?gupath?	m??l?o pay
75	leaf (tree)	Feuille	pum	lif	Oka?y	Okié	Hako	afho	vi?i?
76	root (of tree)	Racine	tutu	rut	N?ka??	n?ka??	Dadi	??ga?	Zèm
77	salt	Sel	guptutu	salt	N?ku?	n?ku?	Manda	?gwa?	m?gahày
78	fat / grease	Graisse	jafututu	fat	Av???	Av???	Dellere	mef???	mb??za
79	hunger	Faim	gatutu	hungri	Zae	Zie	Veelo	n?zi?	Mèt?s
80	hunger	(avoir) faim	ke?	hungri	?? wo?k zae	a? wo?g zie	Nanougo velo	n?zi?leza?	Mèt?s
81	iron (metal)	Fer	fuom	iron	Etyé	Ekie	Djamdi	hanya?	Jamdi
82	one	un (1)	?ie	one	F??k	F??g	Gotel	w?m?'??	Tak?n
83	two	deux (2)	?ie	two	Bae	B?	Didi	mepi?a?	Sarày
84	three	trois (3)	a?a	three	La?	L??	Tati	met??t	màk??
85	four	quatre (4)		four	Nyiin	Nyii	Nai	lekua	Fod
86	five	cinq (5)	mu?	five	Ta?n	Ta?n	Djoi	meta?a	j????
87	six	six (6)	mba	six	Saman	Saman	Djego	ntoko?	K?h
88	seven	sept (7)	te'	seven	Zamgba?l	Zamgba?l	Djedidi	esaambi?a?	Cèsèrày
89	eight	huit (8)	Kpa'	eight	Mwom	Mwom	Djetatati	lefaa?	c??f?? c??f??
90	nine	neuf (9)	te?n	nine	Ebu?	Ebu?l	Djenai	levu'u?	k??raf ka tak?n
91	ten	dix (10)	tu	ten	Awo?m	Awo?m	Sappo	leg??m	gu?
92	come	Venir	saba	cam	?? s??	a? s??	Warougo	lesh?'	Yàha
93	send (tr)	envoyer en commission	fam?	sen	?? lo?m ém??	a? lo?m ém?	Vahugo lelal	ntu?nyi?	zl??n
94	send (tr)	envoyer (qqch à qqn)	ko?vy	sen	?? lo?m (mot djo?m ézi?)	a? lo?m mod dzo?m zi?		ntu?azo?	zl??n
95	walk	marcher	ko??m	waka	?? wulu	a? wulu	Yahou	lezi?	Vaa
96	fall (v)	Tomber	ja	fall	?? ku	a? ku	Dôou	le??m	nd??v
97	to leave, depart	quitter, partir	?ie??	comot / go	?? ko?lo, k?	a? ko?do, a? k?	Sohou Dilou	leg??	l???

(continued)

List ID	List Gloss English	List Gloss French	Bamun	Pidgin English	Bulu	Ewondo	Fufulde	Yemba	Daba
98	fly (v)	Voler	?ie??	fly	?? wu?p	a? wu?b	Firou	lej?	Mb?r
99	pour	Verser	?i?	trowe	?? soe	a? soe	Roufou	lesa?	z??r
100	hit, strike	frapper, donner un coup	ji?gbu	nak	?? bi?bi, é v?? eku?ta	a? bi?mi, a? h?? eku?da		lezap	w??
101	bite	Mordre	jindo	bite	?? koee	a? koge	Gaton	lelo?	Ngàce
102	wash	Laver	jin?y	wash	?? sop	a? sob		n?so?k	P?n
103	wash	laver (habits)	jinbi	wash	?? sop (biyé)	a? so?b (bi?yé)		n?so?kmezoho?	Gòh
104	split (tr)	fendre	jindap	brok	?? sal	a? sal		n?sa?	????àk
105	give	donner	jinum	gi	?? v??	a? h??		leya?	v??l
106	steal	voler (dérober)	jinzu	tif	?? wu?p	a? wu?b	Firou	lej?nyi?	h??l
107	squeeze	presser	jinzu	pres	?? miat	a? miad		?k?	Vaz
108	cultivate	cultiver	jinsen	work	?? be	a? be		n?j?'	Bìz
109	bury	enterrer, ensevelir	Jinfa'	bury	?? be, ?? jam	a? be, a? dzam		letu?	p?s
110	burn (tr)	brûler	jin?uot	born	?? di?	a? di?gi	Woulougo	let?	k??aw
111	eat	manger	jin?e'	chop	?? di?	a? di?	Nyamougo	lepf?t	z?m
112	drink (v)	boire	Jinva'	drink	?? nyu?	a? nyu?	Yekougo	lenu	Sà
113	vomit	Vomir	Jinsi'	vomet	?? yo?	a? yo?	Toutougo	n?za'ne	v??nà
114	suck	Sucer	jinton	suck	?? nya??	a? nya??	Noukougo	lesw?	see?
115	spit	cracher	jin?u'	spit	?? tu?i	a? tu?i	Taiti	n?tso?k	T?f
116	blow (v)	souffler	jinnwo	blow	?? vu??u?lu	a? vu?ndu	Foufougo	n?f?k	n??m?
117	swell (v)	se gonfler, enfler	jin?ar?	high	?? f??m, é vu?s	a? f??m, a? vu?s	Boutougou	??k?'	Fàk
118	bear child	accoucher, donner naissance, engendrer, enfanter	jinnu	born	?? bi?a?y ; é bi?ay m??n	a? bie?, a? bie? m??n	Ndagnougo	n?zihi?	Mbù
119	die	Mourir	Jin?et	die	?? wu?	a? wu?	Magnougo	?kwh?	m??c
120	kill	Tuer		kill	?? we?	a? we?	Mbahougo	n?zwihi	w??l
121	push (v)	pousser	jinfwo	push	?? pu?s	a? pu?s	Yenbougou	letsoo?ne	gò?
122	pull (v)	Tirer	jikwo	drak	?? dutu	a? dudu	Dassougo	n?shu?	gà?
123	sing (song)	chanter	jinve	sing	?? yia	a? yia	Nyamougo	lez?p	Cìk dìrf?
124	sing (song)	chant, chanson	jinkpu	song	Dia?	dia?	NGUIMAL	az?p	Dìrf?
125	play (v)	jouer, s'amuser	ji?o	play	é voe	a? voe, a? b?? bivoe	fIDJOUGO	?g?h?mega	?eewér
126	fear (v)	avoir peur	jinut	fear	?? ma?t?	a? mba?da	Koultanol	lec?'	n??v kw?r
127	fear (v)	avoir peur, craindre	jinsy	fear	?? k??? w???	a? koo? w???		lep?k	Hòvòk

(continued)

List ID	List Gloss English	List Gloss French	Bamun	Pidgin English	Bulu	Ewondo	Fufulde	Yemba	Daba
128	want (v)	vouloir, désirer	jin?wop	wan	?? yi ; é nyam?t?	Yi, a? ko?mbo	Yidougo	?k??	M?y
129	say	Dire	jwopke	tok	?? kat	a? kad	Yetchougo	n?shu?ne	cè?
130	see	Voir	jinkam	see	?? ye?n	a? ye?n	Lahougo	lej??	S?f
131	show	montrer	jindwup	show	?? li?ti	L??d?	Hollougo	n?tswi	d??f
132	hear	entendre	jindwup	hear	?? wo?k	a? vo?golo	Nanougo	n?ju'	Zlìm?
133	know (tr)	connaître, savoir	Jinta'	no	?? y?m, y?m	Y?m	Andougo	n?z?h??	s?n
134	count (v)	compter	jinzie	count	?? la??	a? la??	Limougo	lesa?	n??f
135	be seated	(être) assis	jin?uon	shidon	?? tabe	a? t?b?	Djodougo	lena?a?esi?	???w
136	many	beaucoup (de)	Jinde'	plenti	Abui?	Abui?	Djour	tezaa?	Gòdok?n
137	be white	(être) blanc	jin?u	wite	(?? tab??) efu?mu?lu	(N?), efu?mulu	Danejoum	lepa?	Mà bàren
138	be good	(être) bon	jin?i	gud	(?? tab?) ab??	(T?b?) ab??	Bodoum	lep??	Mà làw
139	be hot	(avoir) chaud	Jinta'	hot	(?? wo?k)	A? wo?g, ayo??	Djaou	lelo	Ma kwàt
140	be hot	(être) chaud	?umansie	hot	ayo??	a? t?b? a?yo??	Wouli	lelote	kwàt
141	heart	Cœur	kwari	heart	(?? tab?) a?yo??	n?n??m	Bennde	at?	n?v
142	be lying down	(être) couché, étendu	jify	sleep	N?l??m	(a? t?b?), m?bu?an, n?yea?n	Walougo	len??	a w?n
143	be standing	(être, se tenir) debout	jibopokeri	Stan	(é tab?) m?bu?a?n, n?yea?n	a? t?b?, a? t??b?) t??t?l??	Danougo	letshi?aj?'	A gàr
144	star	Étoile	jizi	star	(é tab?, é t??b?) t??t??l??	Oti?tie	Kade	mec?'sa?	t?ra
145	liver	Foie	jizi	liva	Otétee	Es??g	Yerine	lepii?	n?nam
146	be cold	(être) froid	lamtum	kol	Es??k	a? (t?b?) a?v??b	Pewol	ef?k	a?i c?c??n
147	knee	Genou	jindie	nee	(é tab?) a?v??b	Ab???	Koppi	atswi'te?khu	t??tib?z
148	be big	(être) grand, gros	jintetutu	bik	Ab???	(a? t?b?) a?ne?n	Manga	lefa?	Ma mordoy
149	be long	(être) long	mut	long	(é tab?) an??n	Aya?b	Djoudoum	lesa?k	Ma jò?
150	mountain	montagne	Fufu'	mountin	Aya?p	N?k??l	Hossene	le?k?'??	wùr?m
151	swim	Nager	fwokwut	swim	N?ko?l	a? dz??g	Nyinougo	lepiika	Cà zlèv zlèv
152	be black	(être) noir	tuku??	blak	?? dj??k	(a? t?b?) évindi	Balejoum	s?h??	Mémìlin
153	be new	(être) nouveau, neuf	jibotami	new	(?? tab?) evi?ndi	(a? t?b?) m?f?fe	Kessoum	eswhi	?àwy?n
154	be small	(être) petit	jniza	smol	(?? tab?) m?fefe	(a? t?b?) a?b??d	Petel	m?	mét??z
155	rock	Pierre	kwot	bik stone	(?? tab?) ti?ts??e?	Ak??g	Tappane	t??	Gùgùn
156	full, filled	(être) plein	Jintamjo'	flop	Ak??k	(a? t?b?) n?dza?la?n	Yewougo	n?do	mewèk,

(continued)

List ID	List Gloss English	List Gloss French	Bamun	Pidgin English	Bulu	Ewondo	Fufulde	Yemba	Daba
157	who?	qui?	jinbosi	wo?	(?? tab?) n?dja?la?n	Za? ?	Moyjo	aw??	Y??
158	what?	quoi?	munfe	weti?	Za? ?	Dzé ?	Doume	Ak??	M??
159	be round	(être) rond	jibomentwi	round	Djé ?	N?	Tekkougo	n?fa?	Mel??ri
160	be red	(être) rouge	tuwuo	red	(?? tab?) an?? ébuma?	ebuma?	Vodjougo	lepa?	Méwizèk
161	be dry	(être) sec	jut	dry	(?? tab?)	(a? t?b?) ev??l?	Yokougo	n?zoo?te	Mekwalkwal
162	seed	grain, semence	w?	sid	(?? tab?) n?ko?t	(a? t?b?) n?ko?d	Aoudi	n?zwi?t	Sèsin
163	sleep (n)	sommeil	k???	slip	N?s??e?	n?so?é	Doydi	lel??k	W?n
164	all	tout, tous	Ji?ga'	all/ ol	Mv??	Oy??	Foumadji	nts?m/?kwa	P?t

(continued)

List ID	DII	Basaa	fe'efe'	mafa	Guidar	Doyayo	Nda'nda'	Kako
1	ya?g	?yo	Nshu	mà	ma? d?fa?	gnayo	Shwi	nùmbù
2	n????	Dîs	Z??h	di	hara d?fa?	Lèllè	n??gh	mìsì
3	da?g	??	Tu?	gid	ki?y d??fa?	Zoulè	tsw??	tó
4	yu?u? hu???	Tjô?	No?kntu?	gotse gid	mu?r ki?y d?fa?	Soussougoo	N?????	mbùrù tó
5	nu??	Liso?	Se'	ldene	zla?y d?fa?	Noungoo	Sw??k	su?
6	m????	Hilemb	La?m	nene	arzl?m d?fa?	Mèllè	Li?p,?i??i??p	?yèm
7	mi?g	Jôl	Nzi?	hetsen	m???k??n d?fa?	Goes	nz??	?yòy
8	t??g	ôô	To?k	ldembad	sl?ma? d?fa?	Tonnè	tu??u?	t??
9	v?g	Ki?	Mi	daya	wula? d?fa?	Dourgo	t??k	?gí?
10	va?a?	Méè	Pe?n	wà	bide ?ak?,	Wallè	P??	???rì
11	nag	Libai li woo	Po?	rì	va? d?fa?	Nannè	pu?	???
12	nag	woo	Mbo?	Yam rì	va? d?fa?	Nannè	pu?	kpé?gélé ???
13	nag f??	Yalak	Kw??? nkap	dzehen	pri?ngi?n	Orgo	ka?ba?	ny???
14	nag f??	Gwalak	Kap	dzehen	pri?ngi?n d?fa?	Orgo	Ka?ba?mbu?	ny??
15	doo	Libal li k??	Khu	Sak	t??l d?fa?	Ndo-yo	?kx??	kol
16	doo	k??	Nkhu	slesheli	sisere	Ndo-ndinra	?kx??	???y
17	doo	b?l	Nkhu	Yam sak	sisere	Ndo-na	?kx??	kíndí
18	vbi?i	Gwèb	Nshunten	Vezi	gabba	Pannè	shu?tu?	m??tà
19	sii	Libùm	Vam	Hud	g?d?f d?fa?	Bou-yo	t??? v??p	m??y
20	ha?g	Libùm	Vam	Hud	g?d?f d?fa?	Bouboloumbo	v??p	m??y
21	kii	Jôb	Te'	Di morov	marba	ngolè	T?t???	?y??bu
22	ya??	Minlà	Nto?	Vayatse	her?e	Yadè	tu?	m??njà
23	m?m	Matjel	Nsi	Pambaz	mbeli?	Gammè	?i??p	m???gìyà
24	zi???	Massai	Nsisi	Kurì	kine?	Metinè	s??s??	m??jàl
25	n????	Héhi	Ncwi?	Vatald	zl??zl??	Ninko	v??gh	yès??
26	fo??o??	Kôô	Mfh?'na?	Ldewed	g?ba d?fa?	Zolè	?g??b??	kònd??
27	kpa?kpa?g	Papai	Mb??	Ldolduk	perzleke	Pekpek-yo	f??	pàp??
28	b?d	Gwalak	Mfh?	Ldolduk	dedeke	mbompto	f??f??	mbùrù
29	?????	Ihon	Nde?'	Durum	m??h??	Goullè	nd??k	làk??
30	za?m?	Nwel	Nkwen	Futur	kuturu	Terri-yo	?ku?	kòndú
31	na?na?n	Mùt	Wenok	Ndo	???f titire?	Dowoo	m??nt??	múm??
32	waye?e	Munlom	Mba?'	Noura	??fa?	wallo	ba??a?	mbám
33	wake??e??	Mudàà	M?nzhwi?e?	Nguza	?ak?	Nisçè	m?zhwi?	nyàrì

(continued)

List ID	DII	Basaa	fe'efe'	mafa	Guidar	Doyayo	Nda'nda'	Kako
34	w????	Nlom	Ndu?	Baki	zile?	wallo	dwi?	mbám gwákí
35	waa	Man, Mangé	Mo?o?	Kra	wi?ina, wank?	waayo	p???kx??	mny?? sìk??
36	ye?e?	Jôl	Ze?n	Ldembad	?lva?	Noungoo	z??	?in??
37	be?	Ngii	Pu?h	Gikile	disi??		ka?k m?? n??m	?yo???
38	vi?d	ù	Th??'	Vad	d?f?a?	Nyemçè	su?'	tu?
39	se?e?	Sôn	M???u?	Kiya	t?la?	Sèyo	ma?nu?	?gw??nd??
40	za??ga?	Hangal	Nam	Patse	pa?ya?	Ottiiyo	n??m	yés??
41	sa?pe?e?m	Mbéébi	Fufu?	Koldar	s?mya? d?da bla?	Zèko	f?? mb??k	pùp??
42	sa?ku?u?d	Hilon	Ndual?	Vani	zlay bu?na?	Goussè	lwi?i?	kùlùtù
43	m??d	Doudoup	Mu'	Mimin	min?e	Metto	m??'	m??l??kà
44	mam	Nob	Mbak	Yam	bu?na?	Bèllè	mb??k	mbìy??
45	ha?g	Isi	C??'	Gid hirkida	d?lva? fu?, delva?	Yepto	ca?'	m??l??tí
46	ha?g	Isi	C??'	Hirkida	delva?, d?lva? fu?	Yepto	ca?'	m??l??tí
47	ha?d sa?ga?	Biték	?wa?'nshi	Wuyak	hoyo?nko	Yollè	sa?k sa?k	m??s??y
48	hi?na?a?	Njél	Nzhi	Chived	tive?	Nbanayo	ma?nz??	njé
49	mam	Malep	Nshi	Yam	i?kile?	Mèmmè	n?s??	m??r??kù
50	zi???	ilép	Vahnshi	Ldoldur	m?mpara?, m?lpa?	Zèyo	n?s??	?ùkù
51	zi???	Ilélèba	Menpii nshi	Ldoldur	m?mpara?nk?	Mèmme-yago	n?s??	?g???
52	lig	Ndàp	Mbee	Ngui	g?l?kk?	Rato	pa?'	tu??
53	Ki?i, lig	Ndàp, yengi	Pe'e? te?nd????	Ngui	da??la	Rato	jia?	tu??
54	vee	Héé	Mo?h	Vogo	o?ffo?	Làyo	mu?'	?it??
55	la?	Tjéé, Bikek	Nshwi?	Sheugeu	s??se?me?	Tein-yondo	?kx??	yo?ta?
56	nim	hiira	Ndhi?mo?h	Vodzok	ene?ne?	Oussè	nz?? mu?'	yì?y??
57	hu??u?m	Libù	Vh??	Hutsad	mo?lo?	Tammè	s?? mu?'	su??it??
58	pag	Iken, ?gwendé	Pi?e?	Gush	g??rdu?, zl??h??	ndango	pi?	k???
59	za?	?köö	Nkh?	Temba	he?lege	Nbanyo	?kx?? ?kx??	k??l
60	ta?a	Likong	Kwe'	Sulum	pi?e	Goyo	?ka?'	k???
61	s??	Gwét	Nco	Gada	??gaya?	Gamssè	ca?k	?ya?mbì
62	p??g	Nùga	N??	Suki Vara	z?ga m?maya	Sobo	n??p	nya?m??
63	w??	Nlamb	Mb??	Ldewed	zluwa?	Nambo	mb??	nja?y
64	ya???a??d	?gwo	Mv????	Keda	k?ra?	Kayoo	mv??	mbìye?
65	ba?l	Njock	S????	Bikine	bekne	Namb-no-yo	s?? nop	nj??ku?
66	?e?e?	Kembêg	Mvi?	Mbuk	haw??	Isyo	zw??	ta????
67	n??g	Inuni	Sa?k	Ndiyak	b??ga	Nosè	s??k ?g??p	n??n

(continued)

List ID	DII	Basaa	fe'efe'	mafa	Guidar	Doyayo	Nda'nda'	Kako
68	kp??g??d	Kùl	Shi?m??nko'	Kuduf	kar?baha?	Goundrè	tsw??m??kw??'	?gu?ru?
69	b??g	?oo	Nu?	Shesheuk	d??h??	Bekçè	nu?	ny?????
70	ze??	Hobi	Mbie?nshi	Kilef	kilfi?	Doutou	fi?si?	nja?nj??
71	go?go?	?ind	Ncat	Tedzi	ten?i?	Ndèdè	shj??	nyin??
72	ba??	Litjéé	Pom	Ldaldi	di?zle	kobarè	mb??k ?g??p	kí?
73	la?	éé	Th??	Wuf	wula?nga?	Tèyo	t??gh	jètì
74	ku??? ko??o	Mbab éé	Kop	Deb wuf	k??zl??pr??	Pokillè	?g??b?? t??gh	kòpù jètì
75	kpag	Hai	Hu	Vah wuf	va? wula?nga?	Npasçè	kx??	mb??rù
76	za?ga?d	Nkan	Nga'	Vezi wuf	su?we?re?	Tè-yesçè	?ga?k	mpéndì
77	ta?m	Bas	Ngwa?'	Bamà	bu?tu?wa?	Mandayo	?gwa??	kwà
78	-	Lihon	Nhu?	Pambaz	s??lh??	Notto	?i?t	kpe??e? mu?t??
79	go?m	Djàl	Nzhie	Mi	ma?ya?	Akillè	?i?ku?'	ja?
80	go?m ya??né	Djàl	Nzhieku?'	Mi a kid	ma?ya?	Akilyakyo	mby?? gh?? ?i?ku?'	wo?k?? ja?
81	h??	Nkong	Te?n??	Zuda	za?mdi	djamdjiyoh	t??	tu?mba?
82	da?gga?	Pog	Nsh?'	Stad	taka?	Mbounou	c??'	w??t??
83	idd??	I baa	P????	Chew	sula?	Érrè	py??	yì?a?
84	ta?a?n??	I aa	Ta?a?	Makar	ho?ku?	Tarrè	te?	yìta?tì
85	ndadd??	I naa	Kw?	Fad	Po?do?	Nasço	kwo?	yìnî
86	n??n?n??	I tan	Ti?	Slam	Zle?	Nonnè	t??	yìtân
87	gu?u?	I samal	Ntoho?	Muko	Zlire?	Nonn-nbounou	to?o?	yìtân j?? w??t??
88	gu?nd?m	I sambog	S??mb????	Tsarad	Bu?hu?l	Nonn-érrè	so?pmby??	yìtân j?? yì?a?
89	ka?andadd??	Djeum	H??	Tsamakad	d??epo?do?	Guèssè	ho?p	yìtân j?? yìta?tì
90	k??gda?ga?	Bôô	V?'??	Chud	Vaita?k	Nanza	v??'??	yìtân j?? yìnî
91	wa??n??o??	Djom	Gha?m/gha?m	Kula	Kla?u?	cobillè	gha?p	ka?m??
92	yal??	Lôô	Ns??'	Sheke	?zza?, ?zzo?	Zamo	mby?? gh??'	njé
93	ba??e?	Hèp iseplà, hep li ké	Nca?h wen o? wu? z?'	Ldende	?zl??n d??fa?	Yamotomsé	mby?? ca?ha?	tòm??
94	ba??e?	Hom, hèp	Nca?h wu? mbi? wen	Ldende	?zl??n s??n z??ga? s??n d??fa?	Yas-mokè	mby?? ca?ha? bi m??nt??	tómsé
95	?ye??mne?	Li ké, hom	Nna?k	Nde	??da?wa?	embomo	mby?? zi?	k??nd??
96	wa??a??l??	Kw??	Nco?	Nkelde	??zb?ra?	Mato	mby?? t??ps??	?a?l??
97	lu?u?l??	Kèè, hébi	Ndo?	Nkelde	?dda?, ?ddo?	Rekmo	mby?? gha?	dìs??, kwa??
98	u?gg??	Puwé	Mfu???	Nmetsere	??v??nge	Ndoukmo	mby?? f??	j??
99	lo?o?l??	Sôp	Nse?'	Ngede	Ubbo?ho?	Goumo	mby?? sa??	wo?s??
100	de?ne?	Bép, hom	Mbu?'	Ntsi	?rga?	Lèmmo	mby?? za?p	m??nd??, s??m??

(continued)

List ID	DII	Basaa	fe'efe'	mafa	Guidar	Doyayo	Nda'nda'	Kako
101	n??nne?	K?g?l	Ndo?m	Ndze	?ha?ra?	Noummo	mby?? cu?gh	kwe?t??
102	wool??	Djowa	Nso?h	Pene	?zla?ya?	Doktmo	mby?? su?gh	we?y??
103	t?l??	Djowa	Nso?h	Pene	?sko??o?	doktombandillé	mby?? su?gh	we?y??
104	ye???e?	Bahàl	Nse?e?	Tsilde	?p??aka?	Seimo	mby?? ba?gh	po?t??
105	p??l??	Ti, Nti	Nh??	Vi	?psa?	Gammo	gha?	ny??
106	li?w k??l??	Nib, wib	Nzh????	Metsere	??h??la?	Pèmmo	zhj??	gu????
107	?????ne?	?r	Nz??'	Detse	????to?	Ndommo	n??'	nya?m??
108	ba?a?le?	Nsèm, Bèl	Nz??'	Peze	Inche?	Nbalmo	cwa? nyi?nya?	???
109	n?m d??nne?	Majona	Nto?k	Ri ndo	??nd??ka?	gnimoo	t???	pu?mb??
110	a?nne?	Ligîs	Nto?o?	Pine	?zla?	Sonmo	cw??	lo??y??
111	la?le?	Jèè	Nz??	Ndiya	?z?ma?	Lemmo	kw??	?ye?
112	z??l??	?ô	Nnu?	Nsiya	?sa?	Yommo	nu?	h???i?ye?
113	be??nne?	L?	Nd??'	Venehe	?para?	Yemmo	l??'	jò
114	a??a?ne?	Nùng	M?mi?	Ntsede	??su?bo?	Yommo	fj??	??????
115	s?d t????ne?	Tolé	Nto?h	Ndzefe	?tu?fo?	Yemmo	ci?gh	sa? m??s??rì
116	u?u?l??	hép	Mfe?h	Shefe	?ffo?ni?	Oulmmo	f??gh	wùbì?ye?
117	h????e?	Hungùl	N?n????	Ngire	??sla?	Hollo	k??'nyi?	wu?w??
118	waa ha?a?ne?	Gwàl	Nzi?	Nya	?wa?	Goulo	z??	
119	n??le?	W?, ?yemb	Nku???	Nmtsi	??mta?	Yorro	fw??	ja?
120	n???le?	Nol, nola	Nzhi?	Nkide	??ba?ya?	Nbommo	zhwi?	gwe?
121	ndagg??	Tiigé, Nugé	Nti?e?'	Nteshe	??re?de?	Yarmo	c??	wo?
122	ta???e?	Hot	Mfa?k	Mete	?da?ha?	Tammo	f??k	pu?s??
123	ye?e? be?e?l??	Tob	Njo?hnshi	Ldi	?ya? g?lya?	Èmmo	gh??p	na?mb??
124	ye?e?	Hémbi	Y??hnshi	Demesh	?ya?, g?lya?	Nàbo	gh??p	je?mb??
125	fo?o?l??	Tùk, mintùk	Ng?? s??	Wali	??sma?ya?	Oksçè	gha?gh ?gxa?	je?mbi
126	nd??gg??	Wo?i	Mbo?h	Gi ldaw	Gezlne?	Hobo	b??gh	ke????
127	nd??gg??	Iseslà	Mbo?h	Gi ldaw	Gezlne?	Nhob-yaki	b??gh	gwe? w??
128	hi?i?l??	Ngôn	Nkwe?'	Nwaya	??gla?, diya? saw?a?	Nhak-yo	k??k	ka?mb??
129	o??l??	Kàl	Nde?n	Gede	?na?ha?	Ndog-yo	s??k	kwa??ye?, g??r??
130	ho??l??	téhè	Nji?i?	Reka	?wla?	Nsék-yoo	yu?	l??p??
131	tu???e?	Héba	Nd??'si?	Tive	?rba?, ?rbo?	ngoun-ngouyo	cw??	??????
132	ki?l??	Nok, noga	Nju?'	Tsina	?pr?ma?	Nken-yo	yu?'	te??ye?
133	ga?a?l??	Yi	Nzhi?	Shune	?s?na?	Nmenè	zh??	wo?k??
134	wa???e?	Songôl	Nsa?h	Ngire	???a?	Nbermo	sha?k	du?w??

(continued)

List ID	DII	Basaa	fe'efe'	mafa	Guidar	Doyayo	Nda'nda'	Kako
135	mba?a?l??	U yii	Nsi? si?e?	Ndiya had	?sso? (tu?go?)	Ndotmo	s?? se?	t??l??
136	???l?? na??a?	?gandak	Y????	Deka	di??a?	Nboltono	y??	?y??
137	de???e?	Mpubi, u Mpop	(B??) f??hta?m	Kodaka	???f b??ba?ra?	nbouyo	f??gh	?u?ja?te?
138	to??o??	Longué	(B??) mbe' ya?a?	Amba	???f mo?ggo?ni?	Bèro	p??k	??? wu?mna?te?
139	ha?a?g	Lèk	Z?' nd??lom	Vazaza	kiti?y k??ka?h??	Lèl-satko	z??' mby?? l??p	??? n?? ?gbe??
140	pi?	wanda	(B??) ndo?m	Zaza	?sso? mo?ggo?ni?	sahsah	mby?? l??p	??? n?? jo??
141	zo??o??	?èm	Mba?'nta?m	Ndav	Ma?kkra?	Ngolè	nti?	??? n?? ny?????
142	u?nne?	U nini	N??ne'	Mana te had	?ssa? ??ffo?	koslè	n??k	te?m??
143	d? yu?ba	Bitété, u téé	Nte?n th??	Ldatsa	?ssa? ziwwe?	Érkilè	si? t??gh	??? k?? ja?k??, na?mba??
144	zoo waa	Hodog	Sa'	Benti	?g??rzla?	Sèya	sa?k	t??m??
145	t??g??m	Libaa	Mbie	Befa	me?m?e?ke?	Silè	bhi?	sís??
146	ho??o??m	nsunè	Fo?h	Talde	kiti?y z??za?ya?	Yariyo	f??gh mby?? ku?	s??k??
147	doo nu?ug	Libôn	Zok	Si Sak	Ta?ppra?	Ndougo	zu?gh	wo?k?? ?u?ku?
148	p??g	Bikeni	M???? ya?a?	Bi	???f metti?gini, m??kk??r??n	Tobilè	wu?'	????
149	l??g	Hâb, hâbag	S?s?? ya?a?	Ndik	???f m??b?a?yi?n	Sibà	sa?gh	??? n?? ????, ??? n?? ???k??
150	g??	Ikoa	Poo	Dza	m??s??rg??	Nkollè	kw??k	??? n?? s??w??
151	peel??	Hôga	Nna?hnshi	Tsi durov	??k??r eke?t i?kile?	Fougmemmè	ly????	ke?tì
152	vi?i/ i??i	Nlanga	Si?si?	Madza	?ssa? titire	Huilo	s??	m??k??
153	?ma???	Nkénd	Fi?	Mbua	?ssa? ma?la?ka?, essa? mele?ke?te?ni	Seggo	b??k	wìnd??
154	?m????	Isii	Me?npi?i? ya?a?	Kute	?ssa? m??ffi?nya, ?ssa? mu?s?u?ko?	Yensçè	kx??gh	j??nja?te?
155	?y??g??	Ngok	Loh	Kwa	ku?:na	Nbeto	lu??u?	mo?n??te?
156	sa?g	Ban	Lo?h	Vavara	?ssa? m?zlna?h??n, ?ba?	Yitigo	wu?t	ta?rí
157	n??n?	Njéé	W?? ?	Ndawa	naway di??	Boma?	w??	to?ndu?na?te?
158	e??no??	Kii-i	K???	Sukime	mi? di??	Noyno?	k??	nda??
159	ki?gi	Limùga	Ka?a?ka?a?	Ndalalak	?ssa? gu?du??gu?lo?		t??	?ge??
160	ye??e??	Koyôp	Pa?'pa'	Nduz	?sso? pepri??ge	Hobbo	pa?k	????????
161	?w??gna?	Nkot	Yo?m	Gora	?sso? m??kl?w?n	nkotlè	yi?p	te?na?te?
162	ye??n?	Bôra	Nzhwia	Tsera	z???ve?	Ninçè	z??gh	so?su?na?te?

(continued)

List ID	DII	Basaa	fe'efe'	mafa	Guidar	Doyayo	Nda'nda'	Kako
163	n?m	Ilo	Lo?h	Dzewen	si?ne?	Yommè	ji?	mbu?m??, kwa?l??
164	?wa?a?pa?d	yosona	Nkwee	Kilek	pa?k?	çesyou	zu?k	ja?k??

Notes: Source: authors. Note: The list of words is extracted from Atlas Linguistique of Cameroon (ALCAM), Number of words: 164. `https://comparalex.org/index.php?page=stdlist&id=3`. the translations have been conducted by Cameroonian linguists.

Table A2: Distance between the Regional Capitals Attended by Students and the Cities with the Highest Proportion of LOI Speakers

Students' School Province	Capital city	Distance to Bamenda	Distance to Buea	Distance to Yaounde	Distance to Douala
BAMBOUTOS	MBOUDA	50.3	249	327	255
BENOUE	GAROUA	826	1108	1110	1172
BOUMBA ET NGOKO	YOKADOUMA	905	917	610	844
DIAMARE	MAROUA	1019	1301	1315	1365
DJA ET LOBO	SANGMELIMA	541	465	169	391
FARO	POLI	721	1080	1046	1031
HAUT NKAM	BAFANG	118	194	295	201
HAUT NYONG	ABONG-MBANG	603	540	237	467
KADEY	BATOURI	721	733	426	660
KOUNG KHI	BANDJOUN	94.5	240	283	247
LEKIE	MONATELE	307	308	81.2	234
LOM ET DJEREM	BERTOUA	632	644	337	571
MAYO BANYO	BANYO	300	528	541	535
MAYO DANAY	YAGOUA	1142	1422	1392	1488
MAYO LOUTI	GUIDER	953	1234	1208	1300
MAYO SAVA	MORA	1026	1307	1371	1372
MAYO TSANAGA	MOKOLO	943	1224	1319	1290
MBAM ET INOUBOU	BAFIA	248	360	130	290
MBAM ET KIM	NTUI	328	390	83	317
MENOUA	DSCHANG	88	168	337	218
MFOUNDI	YAOUNDE	376	307	0	234
MIFI	BAFOUSSAM	78.6	259	301	261
MOUNGO	NKONGSAMBA	155	144	346	151
NDE	BANGANGTE	127	239	252	247
NOUN	FOUMBAN	132	323	336	328
OCEAN	KRIBI	466	245	284	171
SANAGA MARITIME	EDEA	355	134	175	60.5
VALLEE DU NTEM	AMBAM	621	473	249	400
VINA	NGAOUNDERE	598	826	833	1067
WOURI	DOUALA	305	74.8	234	0

Notes: Source: author calculations.