

Where Tools Are Few: Constructing Limited-Word Bilingual Learner's Dictionaries in a Low-Resourced Community in Malawi

Ian David Dicks, *Ciyawo–English Dictionary Project (CEDP), Malawi;
Tabor College, Adelaide, Australia
(iandicks@icloud.com)
(<https://orcid.org/0000-0003-0110-9596>)*

Abstract: Many minority language communities in sub-Saharan Africa are required to navigate national education systems dominated by languages in which they are not largely proficient, and which they lack resources and tools to learn. When taken together with a low political will for resourcing minority languages, these become formidable barriers for learning and development (Matiki 2006: 244, 246; Prinsloo 2017: 20). Many minority languages are also not understood or spoken outside of these communities, and there are few resources to assist people to gain capacity in them, which results in large communication gaps. This paper examines how a team of non-professional Yawo lexicographers in Malawi worked to overcome these barriers in order to produce two limited-word bilingual learner's dictionaries (English–Ciyawo; Ciyawo–English) in book and smartphone application forms. More specifically, it explains how the team was trained, the development of a limited organic corpus from oral interviews as a method for supplementing an unbalanced corpus, and the importance of collaboration. The paper also articulates the linguistic and pedagogical theories that are foundational to the project, including that second-language learners develop second-language competence more efficiently through engaging with a curated headword list of high-frequency and high-relevance words (Nation 2022: 15-17; Bayetto 2018: 12).

Keywords: CIYAWO, POVERTY, BILINGUAL, LEARNER, DICTIONARY, EDUCATION, MALAWI, AFRICA, OXFORD, LEXICOGRAPHY, SMARTPHONE, APPLICATION, CORPUS, LANGUAGE

Opsomming: Waar hulpbronne skaars is: Die maak van beperkte tweetalige aanleerderwoordeboeke in 'n Malawiese gemeenskap met geringe hulpbronne.

Baie minderheidstaalgemeenskappe in sub-Sahara Afrika moet in nasionale onderwysstelsels werk wat oorheers word deur tale waarmee hulle onvertroud is en waarin hulle nie hulpbronne en gereedskap het om te leer nie. Wanneer dit saamval met 'n swak politieke wil om minderheidstale te befonds, word dit gedugte hindernisse vir leer en ontwikkeling (Matiki 2006: 244, 246; Prinsloo 2017: 20). Baie minderheidstale word ook nie buite hierdie gemeenskappe verstaan of gepraat nie, en daar is min hulpbronne om mense te help om taalvaardiger te word, wat tot groot kommunikasiegapings

lei. Hierdie artikel ondersoek hoe 'n span amateur-Yawo-leksikograwe in Malawi gewerk het om hierdie hindernisse te oorkom om twee tweetalige aanleerderwoordeboeke (Engels–Ciyawo; Ciyawo–Engels) met beperkte woorde in boek- en slimfoonformaat te produseer. Meer spesifiek word daar verduidelik hoe die span opgelei is, asook hoe 'n beperkte organiese korpus uit mondelinge onderhoude ontwikkel is as 'n metode om 'n ongebalanseerde korpus aan te vul, en die belangrikheid van samewerking. Die artikel neem ook die taalkundige en pedagogiese teorieë fundamenteel vir die projek onder die loep, insluitend dat tweedetaalleerders tweedetaalbevoegdheid meer doeltreffend ontwikkel deur gebruik te maak van 'n saamgestelde trefwoordlys van hoëfrekwensie- en hoogs relevante woorde (Nation 2022: 15-17; Bayetto 2018: 12).

Slutelwoorde: CIYAWO, ARMOEDE, TWEETALIG, LEERDER, WOORDEBOEK, ONDERWYS, MALAWI, AFRIKA, OXFORD, LEKSIKOGRAFIE, SLIMFOON, TOEPASSING, KORPUS, TAAL

Introduction

The Ciyawo–English Dictionary Project was first conceived in the late 1990's by Australian Baptist intercultural workers working in Malawi who saw the need to communicate with the Yawo in Ciyawo, but struggled to locate adequate tools to help them do so. Although there are strong social, cultural and missiological reasons for wanting to learn Ciyawo, the decision to learn Ciyawo instead of Chichewa, the most dominant language in Malawi, was not an easy decision as there was a dearth of available language learning resources in Ciyawo at that time. The only Ciyawo grammar books and dictionaries in existence were between fifty and one hundred years old and out of print, with only a few collector's versions available in specialty book shops at exorbitant prices. The problem with these tools, however, was more than a lack of availability and cost. They also lacked elements and qualities that would enable second-language (L2) learners to grow their capacity to understand, speak, read and write Ciyawo efficiently.

Initially, to overcome the communication gap, Ciyawo word lists with English equivalents were created and utilised by intercultural workers. These proved to be inadequate as they did not provide definitions for multiple senses or example sentences to show how headwords are naturally used, as well as other useful information for L2 learners. Another problem with the initial Ciyawo word lists were that they were largely translated from English headword lists. These lists represented ideas that native English speakers wanted to talk about rather than the ideas the Yawo had about the world and their life situations.

In 2005, Ian Dicks, an Australian intercultural worker working with Baptist Mission Australia (formerly Global Interaction) and Steffi Nchembe, a Yawo language and culture assistant, began to conceive of developing a Ciyawo learner's dictionary. This dictionary would consist of Ciyawo headwords and other elements that L2 learners of Ciyawo need to grow their capacity to communi-

cate with the Yawo. The notion for a limited-word learner's dictionary was based on research regarding the needs of L2 learners, which shows that L2 learners grow their second language competence most efficiently by learning the high frequency and high relevance words first (Coxhead 2000: 213; Nation 2001: 17-18).

The size of language vocabularies varies greatly. For example, the Oxford English Dictionary (OED) claims to explain the meaning of 500 000 English headwords and phrases (<https://www.oed.com/information/about-the-oed>). There is, however, a vast difference between the number of words in a language and the number of words that the average native speaker knows and uses. Studies show that "native speakers of English typically learn between 15–20 000 word families"¹ (Webb and Nation 2017: 7, 14). It is also recognised that an L2 learner can communicate adequately with a much smaller curated vocabulary. Studies show that high-frequency words are of the greatest value to L2 learners (Webb and Nation 2017: 6). By learning the 2 000 highest-frequency word families in English, as designated by the British National Corpus (BNC), a person has the ability to understand 86–91 percent of discourse and text in a novel, 81–84 percent in a newspaper, 89 percent of a conversation, 90 percent in a film and 89 percent of a lecture (Webb and Nation 2017: 11-12).²

It has also been shown that it is less beneficial for L2 learners to learn additional high-frequency words from the third, fourth, fifth, etc. sets of 1 000 word families than it is to learn additional low-frequency words of high relevance (Nation 2001: 17-18). The Oxford 3 000 and Oxford 5 000 word lists have been developed on this theory. The headwords in the Oxford 3 000 list are selected using two criteria — frequency and relevance for learners (<https://www.oxfordlearnersdictionaries.com/about/wordlists/oxford3000-5000#>). Oxford University Press (OUP) identified the highest frequency words, drawing on the BNC and the OED corpus (Shaffer 2005: 136). To this high-frequency list they added words of 'high relevance' to English language learners by selecting high-frequency words from "a specially created corpus of Secondary and Adult courses published by Oxford University Press" (<https://www.oxfordlearnersdictionaries.com/about/wordlists/oxford3000-5000#>).

Dicks and Nchembe saw the value of learning high-frequency and high-relevance words for L2 learners of Ciyawo. Initially, this interest was in the concept of developing a high-frequency and high-relevance word list in Ciyawo for L2 learners of Ciyawo. However, through research it became obvious that there were not adequate resources available to do this. There was neither a Ciyawo general corpus or a high-relevance corpus from which to develop such a list, nor were there human resources to utilise the data if such lists emerged.

During exploration of the value of high-frequency and high-relevance words for L2 Ciyawo learners, Dicks and Nchembe became acutely aware of the Yawo community's language learning context. They noted the lack of resources and tools to help Yawo students learn English as a second language, which at the time was the language medium of instruction for upper-primary, secondary and tertiary education in Malawi.

The Yawo education experience

The Yawo of Malawi are disadvantaged within the Malawian education system, as they are constantly playing catch-up throughout their educational experience. This is due to Chichewa being used as the primary medium of communication in primary school. The Yawo are also disadvantaged when it comes to engaging with government social services in the areas of health, law and order, civil society and development initiatives, as well as religious activities. The majority of professionals who come to work in Yawo-dominant areas are not proficient in Ciyawo, nor are they able to grow their capacity in the language even if they wanted to, as there are few resources available to help them.

This situation is not unique to the Yawo. It is also true of many minority language communities in Africa and around the world. Trudgen (2000: 68) says that in many minority language communities there is "a crisis in being understood". What Trudgen (2000) says of the Yolŋu people, an Aboriginal community in Northern Australia, reflects in many ways the reality of the Yawo and other minority language communities throughout southern Africa. The Yawo sit in meetings, listening to Chichewa and/or English, but often do not fully understand these languages. Alternatively, they may understand but do not feel competent to engage in the dominant language being spoken and so they remain quiet. This results in a tangible isolation in their own country. Trudgen (2000: 69) similarly points out:

Poor communication stops people receiving almost all news or knowledge from outside their language and cultural domain. This includes day-to-day news and general information. It also includes what may well be life-saving information from health professionals. It stops them knowing what they are giving consent for, how to comply with medical instructions and how to intervene in their own health problems. In this way, poor communication directly impacts on high mortality rates.

In Malawi it is taken for granted that Chichewa is universally understood; however, this is not the reality. For example, it is common for the Yawo to return from a visit to the hospital not fully knowing their diagnosis or the treatment regimen. This is not because it was not explained, but because there is often a communication gap. The professional did not understand or speak even a modicum of Ciyawo and the patient was not conversant in English, or fully conversant or confident enough in Chichewa to ask questions.

Trudgen (2000: 77) states that the communication crisis operates two ways: firstly, professionals are unable to communicate with people in minority language communities about matters that impact their lives and well-being, even when discussing basic concepts using basic terms, and secondly, people in minority language communities struggle to share their needs and ideas with professionals in a dominant language.

The Yawo context in Malawi

There are 17 languages spoken in Malawi (Matiki 2001: 203). Four languages are considered official languages, namely Chichewa, Citumbuka, Chitonga and Ciyawo, together with English (<https://www.immigration.gov.mw/citizenship/list-of-languages-in-malawi/>). Chichewa and English were declared the official languages of Malawi in 1968, with 50,2 percent of the population claiming to be first-language (L1) speakers of Chichewa and 4,9 percent claiming to understand English at that time (Matiki 2001: 201). According to Matiki (2001: 215), English became the "instrumental, regulative, interpersonal and imaginative/innovative language" as it was the only language accepted in education, judiciary, parliament and the media. However, due to English not being widely understood or spoken, "Chichewa acts as the de facto national language and English is the de facto official language in the country" (Reilly 2019: 32).

Although it has been widely reported that sub-Saharan languages are not in danger of being supplanted by languages of the colonisers, they are at risk of being supplanted by other dominant languages in their countries. Brenzinger (in Kandybowicz and Torrence 2017: 2) says, "the most immediate threat to minority African languages are posed by other local languages or sub-national languages", which is the case with Chichewa in relation to the other languages in Malawi.

Ciyawo is a widely spoken language in South Eastern Africa. It is the first language of more than 3,2 million people in Malawi, Mozambique and Tanzania (Baldauf and Kaplan 2004: 85, 155). In Malawi, more than 2,3 million people identify as Yawo and they make up 13,3% of the population (Malawi Data Portal 2018).

Malawi, however, is one of the least developed countries in the world with half of the population living below the poverty line when measured against the Multidimensional Poverty Index, which assesses people's lives across four equal dimensions of Health and Population, Education, Environment, and Work.³ Significantly, under the Education dimension, low levels of literacy and issues with schooling contribute to the nationwide level of multidimensional poverty in the country. However, other indicators reveal the depth and breadth of poverty for the Yawo, as the incidence of multidimensional poverty is 65,7 percent for people in rural areas, which is where the Yawo predominantly live, compared to 20 percent in urban areas (National Statistical Office Malawi 2022: vi). Moreover, children and youth are among those most impacted by poverty in Malawi — 63,8 percent of children between 0 to 9 years and 61,8 percent of children aged 10 to 19 years record the highest incidence of poverty in the country (National Statistical Office Malawi 2022: vi). The significance of these statistics is amplified at a district level as Mangochi (78,4 percent) and Machinga (78,2 percent) are the districts in Malawi that have recorded the highest percentage of people experiencing multidimensional poverty. These are areas in which the Yawo are the majority ethnolinguistic community (National Statistical Office Malawi 2022: vi).

This data indicates that Malawi is suffering an education crisis. This is supported by other data that shows that the percentage of children unable to read at a level appropriate for their age is estimated to be 89 percent in Malawi (UNESCO 2024). Successive governments have attempted to address the education issue in various ways. For example, tuition fees were abolished in 1994 for all government primary school education. In 2018, the government also announced its intention to abolish fees for government-provided secondary education.

However, the education crisis in Malawi is more nuanced than merely poor literacy and numeracy skills, and truancy. The education system has largely benefited first language Chichewa speakers since 1968 when English and Chichewa were adopted as the national languages of Malawi (Trudell 2016: 44). In 1969, Chichewa was established as the medium of instruction for the first four years of primary school and English became the language of instruction from Standards 5–8 in primary school and throughout secondary school (Matiki 2001: 206). English is also a mandatory subject in which a pass mark is required for the Primary School Leaving Certificate examinations in Standard 8; the Junior Certificate examinations in Form 2, and the Malawi School Certificate of Education (MSCE) in Form 4 (Matiki 2001: 206).

In 1996, once multiparty democracy was established, the education policy was changed to recognise and structurally include other languages in the educational life of Malawi. It was hoped that this would lead to a new dawn in education for the speakers of minority languages. A new education policy was written that directed mother tongue languages to be used as the medium of instruction in the first four years of education (Reilly 2019: 33). English would continue to be used as the medium for instruction from Standard 5 onwards. Apart from a few pilot programmes, however, the Ministry of Education did not implement this policy in any significant way, upholding the status quo.

Although people who are ethnically Chewa only make up 40 percent of the Malawian population, Chichewa has become the *de facto* first language of several other large tribes in Malawi, including the Ngoni and the Lomwe. When all first language speakers of Chichewa are counted together, they represent approximately 70% of the population of Malawi, which is a large and influential block (Matiki 2006: 250).

In 2014, the language policy for education changed again. This time it stipulated that the language medium of instruction would be English from Standard 1 in primary school through to Form 4 in secondary school (Trudell 2016: 45). This policy was meant to bring more equity to the education system, however it has also not been widely supported or implemented.

There are several reasons for the lack of implementation of the minority language policy as well as other minority language initiatives, including the social significance of Chichewa in Malawi. Chichewa serves an interpersonal function and operates as a link between people with different cultural and linguistic backgrounds in Malawi (Matiki 2001: 212). According to Matiki (2001: 212), Malawians are more likely to use Chichewa than any other language to bridge linguistic

differences, including in the education system. Issues of low language competence among teachers in minority languages and English is also a factor, as is linguistic centrism within the Ministry of Education and District Education Departments, and the overall paucity of educational resources for the population (Matiki 2006: 246-247; Trudell 2016: 39). Malawi does not have sufficient resources for its current education system, let alone a "translanguaging" approach to education (Kamwendo 2000: 6; Reilly 2019: 34). The extent of the education crisis was made evident to the project team when visiting 60 primary and secondary schools in the Mangochi District during 2024 to conduct seminars on how to use a dictionary and dictionary smartphone application (app). One primary school on the Eastern shore of Lake Malawi had 1 363 students and 8 teachers, resulting in a student-to-teacher ratio of 170:1. Moreover, the high number of students means that in some schools, classes are conducted in the open, under trees, as there are insufficient classroom blocks. In regard to the paucity of language resources, at one large primary school in the Mangochi District the project team was shown a small closet containing approximately 20 books, which is vastly inadequate for a primary school of more than 2 000 students. The consequences of this are seen in the Malawi 2018 Census, which indicates that the percentage of Yawo completing primary, secondary and tertiary education is lower than any other ethnic group in Malawi (National Statistical Office Malawi 2019).

The impetus for a wider project

It was from this understanding of the Yawo education context that the Ciyawo–English Dictionary Project (CEDP) was created as a vehicle for developing two bilingual dictionaries for two different end-user groups. One dictionary, the Ciyawo–English Learner's Dictionary (CELD), would be developed primarily for L1 English speakers who want to learn Ciyawo for communicating with the Yawo in their language, thereby promoting dignity and equity in dialogue and conversation (Dicks 2022; 2024a). The primary end users of the CELD are non-Yawo, English-speaking adult learners. The other dictionary, the English–Ciyawo Learner's Dictionary (ECLD), would be developed for L1 Ciyawo speakers, primarily students who need to develop their competence in English for educational and employment purposes (Dicks 2018; 2024b). The target audience for the ECLD would be students in Standards 6–8 of primary school and Forms 1–2 of secondary school in Malawi.

The need for both of these dictionaries was seen as important and urgent. However, since the project team lacked adequate resources with which to develop the CELD — namely, a word list of high-frequency and high-relevance headwords in Ciyawo developed from a corpus, as well as the identification of key word senses and natural example sentences — the team decided to embark on developing the ECLD first, as resources were available to begin this part of the project.

Methods of development

It is well documented that certain resources are deemed essential for the construction of modern, bilingual learner's dictionaries. At centre stage is the corpus, consisting of machine-readable written texts for determining headword lists, sense hierarchies, definitions and example sentences that are to be included in the dictionary articles (Atkins and Rundell 2008: 53). However, even the best and largest corpus is of little use without people with educational, linguistic, lexicographic and technological expertise to design a dictionary for specific end users, and to construct it by utilising the data appropriately. Moreover, anyone who has been involved in the management of a dictionary project knows that they are multi-year, even decades-long projects that require significant funding for equipment, software, rent, salaries, consultancy, publishing, etc. A project is also dependent on oversight and government and community support to ensure that the end product is constructed to an acceptable standard and aligned with the ideals and needs of the community.

In the Yawo context of southern Malawi, virtually all of these resources were unavailable in any significant manner at the conception of the project, except for community support. Prior to 2005, there was not a recognised standard orthography for Ciyawo (Centre for Language Studies (CLS) 2005), and up until 2005, Ciyawo was written idiosyncratically in Malawi and other countries. Even after a cross-border initiative to standardise the Ciyawo orthography involving linguists from Malawi, Mozambique and Tanzania, some linguistic and education departments continued to write Ciyawo idiosyncratically, a decision that hampers the development of literature in Ciyawo to the present day (Banda et al. 2008).

A formidable barrier to creating a high-frequency Ciyawo word list was the dearth of machine-readable text available in Ciyawo with which to develop a Ciyawo corpus. The available texts in Ciyawo were more than a century old. One of these was Yohanne Abdallah's (1973) historical account of the Yawo, written in Ciyawo in 1909. There were also several collections of Yawo-sacred stories and proverbial tales documented and translated by Duff Macdonald dating from 1881 and 1882 (Macdonald 1881; 1882). The only recent texts of significance were a Yawo Bible translation produced by the Bible Society of Malawi, and portions of translated Scripture and Scripture teachings that were being produced by translation teams in Malawi and Mozambique in the late 1990's and early 2000's.⁴ The only other recent text was a collection of traditional Yawo proverbs and proverbial stories, *The Wisdom of the Yawo*, published by Ian Dicks in 2006 (Dicks 2006).

Also available, but not usable for corpus construction were Yawo word lists and older dictionaries, including Edward Steere's Collections for a *Handbook of the Yao Language* (1871), Alexander Heatherwick's *A Handbook of the Yao Language* (1902) and Meredith Sanderson's *A Dictionary of the Yao Language* (1954). Apart from not being in a machine-readable state, none of these dictionaries were constructed from a corpus and they contained archaic terms and definitions

that are no longer commonly used by the Yawo. These texts, however, would still be useful as refining tools once a corpus was developed, as would the *Mgopolela Malowe Jwa Ciyawo* (CLS 2013), a monolingual Ciyawo dictionary.⁵

Aware that the previous dictionaries were dated and largely developed without a corpus, the project team decided to do everything possible to produce a general corpus in Ciyawo from which to construct the CELD. However, there is much debate regarding what constitutes a valid corpus (De Schryver and Prinsloo 2000: 91). A corpus is commonly defined as,

a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research (Sinclair 2004).

Moreover, "a general corpus is typically designed to be balanced, by containing texts from different genres and domains including spoken and written, private and public" (De Schryver and Prinsloo 2000: 91).

Even the task of producing a general corpus was a challenge for an African language with so little available written text. Moreover, the written text that was available was largely limited to a single semantic domain — Christian Scriptures. Nevertheless, it has been suggested that "any corpus — however unbalanced — is to be a source of information and indeed inspiration. Knowing that your corpus is unbalanced is what counts" (Atkins et al. 1991). The unbalanced nature of the Yawo corpus is one fact that the project team were fully aware of.

The aim then became to create what De Schryver and Prinsloo (2000: 92) call a "structured corpus", which is the first step towards creating an organic corpus. Available machine-readable written texts in Ciyawo were gathered, mainly in the domain of Christian Scriptures and teachings. To these were added additional transcribed texts from recorded oral interviews in a variety of semantic domains. The recorded interviews were conducted in domains including but not limited to the following: (i) work (well digging, farming, food preparation, fishing, grafting, business, harvesting, hunting, piece work, making local fermented drinks, clay pot making, carpentry, preparing groundnuts, making hoes, making metal pots); (ii) travelling; (iii) family life (marriage, traditional rules for pregnancy, food, cooking); (iv) traditional medicine (role of a healer, use of charms); (v) Yawo traditional religion (practices and beliefs, sorcery, witchcraft); (vi) Islam (practices and beliefs, religious differences); (vii) historical events (war in Mozambique); (viii) health and well being; (ix) Yawo Games; (x) initiation (boys, girls, pregnancy, chieftaincy songs); (xi) stories (traditional Yawo tales, histories); and (xii) death (funerals).

The bank of text consisting of Scripture and Scripture teachings were mainly in the narrative genre and constituted 448 199 words of the corpus. In regard to the oral interviews, more than 190 interviews were conducted, adding more than 322 000 words to the corpus. The total number of words from all sources in the Ciyawo corpus came to 770 871.⁶

The corpus, however, was unbalanced, as 60 percent was made up of text from Scripture and Scripture teachings and 40 percent from transcribed oral interviews. Some people have raised questions regarding the unreliability of an unbalanced corpus, as well as the unreliability of a corpus developed from spoken data, especially unscripted conversations (cf. De Schryver and Prinsloo 2000: 94). However, it is "any corpus compiler's task to attempt to assemble a representative corpus for his/her specific needs" (De Schryver and Prinsloo 2000: 92). The specific need identified by the project team was to identify the 3 000 high-frequency and high-relevance words in Ciyawo, and to define them for an English-speaking audience in the CELD. Considering this, the oral nature of a large portion of the source material is seen as a strength, since it captures speech that is natural and frequent, and not at a higher and more rarified literary level. Ultimately, corpus design is constrained by what is available, and if there is no written text with which to create a more so-called balanced corpus, the compilers must rely on whatever means are available to them.

It has also been argued that different genres of speech events are of different value to a corpus, with unscripted conversations at one end and monologues and lectures at the other (De Schryver and Prinsloo 2000: 94). The project team used what is known as an "informal interview" approach when conducting oral interviews on selected domains, which placed the genre of speech at the monologic end of the spectrum. The people interviewed were those whom the community identified as having knowledge of a specific topic. The interviewers used various questioning methods to elicit extended responses, including the grand tour question, which allows a person to give an overview of the topic in response. This type of question was used predominantly because it encouraged a conversational tone and allowed participants to talk about what they considered important in regard to the topic, rather than focusing on what the interviewer deemed significant (Fetterman 2010: 40).

Once the interviewing and transcribing was complete, a corpus query tool, Wordsmith Tools 6.0, was used to process the text files and provide statistical information about the database. It also gave concordance lines showing how words are used in context, and enabled the tagging and assigning of word classes. Knowing the unbalanced nature of the corpus required the team to diligently sort, demote and remove lemmas that were overly represented, such as religious terminology. While Wordsmith works well as a query tool, it does not help in identifying low-frequency but high-relevance words. For this, the team consulted privately produced word lists and monolingual dictionaries, including the *Mgopolela Malowe Jwa Ciyawo* (CLS 2013), as well as older Ciyawo Dictionaries mentioned earlier, and the Yawo community.

The need for developing two dictionaries

The project was ambitious from the start. The aim was to produce two dictionaries for two different end-user groups. Some individuals inquired as to why

one bilingual dictionary could not be produced and then reversed or back-translated for the second language community. The reason this is not possible is because bilingual dictionaries are not like eggs — they do not flip easily, especially in the case of Ciyawo and English. This is because the high-frequency word lists are different for both groups of end users. However, even when actions or objects have gloss equivalents in both languages, the meanings applied to them can be different for each "languaculture" community (Agar 2002: 60). Even when the reality being defined seems to be ubiquitous to both language and culture communities — such as *bicycle/njinga*, *dog/mbwa*, *farmer/mlimi*, *to pray/kuswali* — it was found that L1 speakers of Ciyawo and English apply different meanings to these realities. This means that appropriate headword lists are required for each languaculture community, applying L1 definitions to each sense so that L2 learners can grow their understanding to include the L1 meaning of the headword rather than their own L2 cultural meaning.

The English–Ciyawo Dictionary

The purpose of the English–Ciyawo Learner's Dictionary (ECLD) is to assist L2 Ciyawo speakers of English in growing their English proficiency so they can navigate the Malawian education system, gain employment, and interact with L1 and L2 English speakers internationally.

Therefore, the headword list for the ECLD needed to consist of the 3 000 most high-frequency and high-relevance words for L2 learners of English. The project team did not have the ability to develop such a list. OUP was approached and asked if their 3 000 word list could be used, as well as definitions up to four senses, and example English sentences for these senses. While the agreement took time to arrange, OUP was helpful, generous, and accommodating throughout the entire process of interaction. Permission was granted to translate the OUP definitions into Ciyawo. For this task, a meaning-based method of translation was used. This method differs from a literal translation method in that it aims in "retelling, as exactly as possible, the meaning of the original message in a way that is natural in the language into which the translation is being made" (Barnwell 1986: 9).

The target audience for the ECLD is upper-primary and lower-secondary school students, unlike the *Oxford Advanced Learner's Dictionary* (OALD), which has been primarily designed for adult learners. This meant that a number of example sentences in the OALD 8th edition needed to be changed, as the ideas expressed in them were not easily recognisable by the Yawo end users (Hornby 2010).

To assist learners further, the translators sought to use only the 3 000 key English headwords in the example sentences. This enables Yawo learners to look up the words used in the English example sentences that they do not understand. Unfortunately, the OALD does not adhere to this principle strictly and includes words in a number of example sentences that are outside of the 3 000 key word list. To overcome this problem, permission was sought to use a consulting editor to locate replacement example sentences for all the senses

missing appropriate example sentences. This included senses in which the example sentences expressed an idea uncommon in a Yawo context, example sentences that used words outside of the Oxford 3 000 word list, and instances where an example sentence was not provided.⁷ Overall, the consulting editor provided 1 386 appropriate example sentences.

Training non-professional lexicographers

At the beginning, the project lacked people with the skills and abilities required to produce a dictionary, in particular lexicographers. Lacking available expertise, Dicks and Nchembe set to work learning the art and craft of dictionary construction. Dicks attended an intensive introductory course on lexicography at Stellenbosch University. Following this, Dicks and Nchembe researched practical lexicography using the works of Landau (2001), and Atkins and Rundell (2008). They also attended AFRILEX conferences to learn from people who were working on other dictionaries in the African context.

Drawing on this learning, research, and personal experience as learners of Ciyawo and English, Dicks and Nchembe created a style manual and included elements in the dictionary article that would be most helpful for L2 learners of both languages.

The dictionary articles in both dictionaries include the following elements:

- (1) Headword: The word that an L2 learner would be searching for.
- (2) Senses: This includes up to four key senses of meaning in the ECLD and up to three in the CELD.
- (3) Translation equivalent in L1 of learner: This applies if a lexical equivalent exists.
- (4) Definition: The definition of the headword in the learner's L1.
- (5) Example sentence in L2 of learner: This sentence or phrase is meant to show how the headword is used naturally according to the sense.
- (6) Translation of example sentence in L1 of learner: This would include the translation equivalent if a lexical equivalent exists.
- (7) Other tense forms: Verbs are shown in three tenses, namely present, continuous and past tense.
- (8) Other word forms: Other forms of the headword are shown if the word is an adjective.
- (9) Plurals: The plural form of a noun is given.
- (10) Part of speech: Ciyawo grammar terms are used in the ECLD, such as *lina* (noun), *msali* (verb), *mlondecesya* (adjective), *mjonjecesi* (adverb), *mpecesi* (preposition), etc. English grammar terms are used for parts of speech in the CELD.

After receiving feedback from end users on the ECLD that was published in 2018, several changes were made to the dictionary article for the construction of the CELD. A translation of the Ciyawo definition in English was added so that learn-

ers could read the definition in two languages. The number of senses was also reduced in the CELD from four to three key senses.

The lexicographical team

The original Yawo lexicographers were chosen primarily for their language capacity in both Ciyawo and English. They were also required to have completed the Malawi School Certificate of Education (MSCE), which is a recognised competence marker in English. Included in the first group of lexicographers and field testers employed were several primary and secondary school teachers who provided insight into the end users of the ECLD.

All of the lexicographers were provided with initial training, which included teaching on the nature of a dictionary, different types of dictionaries, the hallmarks of a learner's dictionary, the importance of understanding the end user: including their needs, abilities and their perspectives, the objectives of the Ciyawo–English Dictionary Project, and most importantly the principles and methods of defining words. This teaching covered various styles of defining words, including genus and differentia, synonym, typical, and complete sentence styles, as described in several practical guides to lexicography.

A key aspect of a learner's dictionary is the provision of an example sentence in which the headword is used appropriately according to the sense of the word. The ECLD used primarily example sentences from the OALD 8th edition and from the consulting editor, as already mentioned. However, for the construction of the CELD, the newly created Ciyawo corpus was the main source of these example sentences. It was not always possible to use the corpus, since available sentences and phrases were not always appropriate or clear. For the senses without appropriate example sentences, recorded role-plays were conducted by the team to capture the way a word is used in natural dialogue. This method was found to be helpful and also proved to be a lot of fun.

Collaboration

In a low-resource community collaboration is essential — not just for dictionary construction, but in every aspect of life. The Yawo realise this and have many proverbs that talk about the necessity of working together to achieve a purpose, such as 'Mtwe umpepe wangatwicila cipagala' ('One head does not carry a small thatched roof') (Dicks 2006: 96).

In many ways collaboration is what made it possible to undertake and complete this dictionary project. From the very beginning, collaboration was required between L1 English and L1 Ciyawo speakers, as neither community had the language abilities to construct a bilingual dictionary on their own. Collaboration extended further to virtually every aspect of the project. Within the lexicography office, people operated collaboratively: Knowing that individually people

were under-skilled, the lexicographical team always worked in pairs when creating and reviewing dictionary articles, and role-plays were conducted as a group.

Knowing that the team did not have the skills required for designing and constructing a single dictionary, let alone two, collaboration was sought with OUP, which resulted in permission to use designated parts of the OALD 8th edition. OUP is a large multinational company that could have declined these requests. Instead, to their credit, OUP helped this small project team to achieve its goals, while also protecting their intellectual property and brand.

The team collaborated with the CLS at the University of Malawi, Zomba. The CLS provided education and training on the Ciyawo Orthography. Furthermore, two Yawo-speaking academics from the CLS became consulting editors on the project. They contributed by reviewing and providing editorial insights on both dictionaries. The CLS also helped locate several Yawo graduates in linguistics who joined the CEDP lexicographical team.

Collaboration also occurred with the wider Yawo community. Yawo-speaking school teachers and school principals helped to facilitate field testing with students and teachers in primary and secondary schools. The Paramount Yawo chief, Ce Kawinga, visited the project on several occasions, as did other senior chiefs and local village headmen and women to understand the project and give encouragement. Senior Yawo chiefs also sent representatives for training as reviewers, and gave the team feedback and approval to publish the finished work.

The Malawi Ministry of Education provided support. This came after publishing the first English–Ciyawo Learner's Dictionary in 2018. The Minister of Education at the time, Hon. Bright Msaka SC, was the keynote speaker at the launch of the ECLD in Mangochi, and encouraged parents to purchase the dictionary for their children and the Ministry of Education to support the initiative. The recommendation of the Minister of Education facilitated collaboration with the Ministry of Education in Mangochi, enabling the third phase of the project to move ahead, which was to conduct seminars in primary and secondary schools in the Mangochi District on 'How to Use a Dictionary and Dictionary Smartphone Application'.

Project learnings

There were many lessons learnt from this project. The first is that dictionary construction takes longer than anticipated. The ECLD took 12 years from conception to publication, and an additional 4 years to complete and publish the CELD. Thereafter, it took an additional 2 years to release them as smartphone apps, add sound files, and conduct sensitisation workshops in primary and secondary schools in the Mangochi District on how to use them.

Another learning relates to the difference between the needs of dictionary end users and their ability to purchase one. Although Yawo students wanted and needed to learn English, very few had the financial means to do so even at a sub-

sidised price. The ECLD was published and printed in book versions in 2018 and the CELD in 2022. The initial print runs were small: Only 2 000 copies of the ECLD were printed, and of these 1 500 copies were distributed free of charge to primary and secondary schools in Yawo-speaking areas of Malawi, while only 500 copies were sold to the general public. For the CELD, 1 000 copies were printed, of which 512 copies were distributed without cost to schools and 488 copies were sold through various bookstores in Southern Malawi for between US\$5–\$10. Although the demand for the dictionary was great, the ability of individuals and the Malawian Government to purchase them was extremely low. Due to the poor economic circumstances of many Yawo and Malawians, books and dictionaries are rarely purchased because of their cost.

Dictionary smartphone apps

With the introduction of smartphones in Malawi and the high adoption of them by many households, even among the Yawo, it was decided to investigate publishing the ECLD and CELD in smartphone apps that could be operated without the need for cellular data once the dictionaries were downloaded. The 2018 census indicates that 51,7 percent of Malawian households own a mobile phone. Significant for this project, 47,1 percent of households in the Mangochi District own a cellular phone, which is the fifth highest level of cellular phone ownership of the 13 districts in Southern Malawi, apart from cities (National Statistical Office Malawi 2019). Moreover, the number of cellular connections continues to rise. In 2025 there are 13,2 million cellular connections, which represents 60,3 percent of the general population (Datareportal 2025).

Once again, the team lacked the technical expertise or financial resources to undertake such a venture independently. The idea of creating smartphone apps was initially discussed with OUP; however, their suggested pathway was technically and financially out of reach. To find a solution, the team investigated the Dictionary App Builder (DAB) programme, developed by the Summer Institute of Linguistics (SIL), an international Scripture translation organisation. The DAB enables non-professional app developers equipped with an appropriate lexicon data file to build a dictionary app at an extremely low cost. The two CEDP dictionaries were constructed using the dictionary writing system TshwaneLex, which provides multiple options for exporting a lexicon database. The DAB is a sophisticated programme that enables the construction of a dictionary smartphone app with all elements of a dictionary article that are beneficial to L2 learners.

The completed dictionary apps have several advantages over the paper form of the dictionaries. The app enables the end user to search for words in multiple ways, including without knowing how to spell the word fully. Morris (2021: 36) highlights this as a benefit, saying that "electronic dictionaries are structured in a completely different way and no longer rely on the alphabetical access structure as the only access, the outer alphabetical access structure loses its status as the default access structure". The DAB app also allows the end user to search

for words that are only partially known, and to search for and find words that are misrepresented orthographically, which according to Morris (2021: 39) is the sign of a good electronic dictionary interface.

The DAB app also displays each article on a single screen, unlike an e-reader version of a dictionary. This reduces the overpowering affect that a page of articles can have on a novice. The DAB app also allows the dictionary to have an audible dimension, which greatly adds to its pedagogical function. In the second version of the ECLD and CELD apps, sound files were added so that end users can hear the headwords spoken by L1 speakers of Ciyawo and English, which assists with their speech reproduction. Once again, the construction of the apps was made possible through collaboration with the SIL who provided the DAB programme free of charge. The only costs associated with publishing the dictionaries as smartphone apps, have been paying a technology consultant, who has directed and assisted the building of the apps for Android and iOS. Both dictionaries are now available on the Google Play Store and the Apple App Store free of charge.

Finally, the Ciyawo–English Dictionary Project has been a collaboration between the Dictionary making team and financial supporters, who were predominantly ordinary people in Baptist churches of Australia who wanted the Yawo to have more opportunities to complete primary, secondary and tertiary education. My own role, as senior editor and project manager has always been part-time and has been supported largely through people in Australian Baptist churches. Baptist Mission Australia have supported this project and overseen the finances and tax deductible giving, which has enabled this project to be completed.⁸

The cost of this project is hard to establish, as the costs associated with the senior editor and project manager role were not accounted for between 2008–2016. These roles were part-time and conducted in conjunction with other roles and responsibilities in Malawi. Overall, the senior editor and project manager roles were undertaken approximately two days a week for 14 years, and then one day a week for another 4 years.

It is estimated that USD \$385 000 was spent to develop, publish and print two dictionaries in book form, as well as to produce smartphone apps for both dictionaries so that they could become available on Android and iOS.⁹

Conclusion

Many minority language communities in sub-Saharan Africa are under pressure of becoming redundant as mediums of communication, as they are undervalued and under-resourced in language learning tools. The result is seen in low education outcomes, inadequate delivery of social services, a sense of disempowerment, and isolation. What has been shown in this article is that a team of non-professional lexicographers can produce language learning tools through collaborative

effort. There are many other minority language communities in similar situations to the Yawo. Realistically, dictionaries for these communities are unlikely to be profitable for large dictionary corporations in the short term. However, collaboration with minority communities will greatly benefit communities, making learning more equitable and assisting in closing the communication gap.

Endnotes

1. "A word family consists of a headword (for example, 'assume'), its inflects ('assumes', 'assumed', 'assuming'), and its derivations ('unassuming', 'unassumingly')" (Webb and Nation 2017: 7, 14).
2. Others have suggested that a better threshold for an L2 learner is 3 000 highest-frequency word families as this would increase their ability to understand 98 percent of most reading materials and 95 percent of vocabulary used in spoken discourse. Dang and Webb (2016) have suggested that high-frequency words lists should count lemmas and lemma headwords instead of word families as "lemma headwords are the most commonly used unit of counting both inside and outside the classroom" (Webb and Nation 2017: 11-12). Either way, there is agreement that learning high-frequency words is key to closing the communication gap for L2 learners.
3. There are 13 subcategories, referred to as indicators, under these four headings: Food Security, Drinking Water, Nutrition, Sanitation, School Attendance, Literacy and Schooling, Asset Ownership, Housing, Rubbish Disposal, Electricity, Child Labour, Job Diversity, Unemployment (National Statistical Office Malawi 2022: 7).
4. There were some other texts that were very old, and that were not accessible to be transcribed. The most significant texts in the Ciyawo language were mainly other Bible translations, including several New Testament books translated by Alexander Hetherwick in the 1890's, including *Utenga Wambone Wa Luka*, *Masengo Ga Wandumitume*, *Utenga Wambone Wa Marko*, *Utenga Wambone Wa Matayo*, *Utenga Wambone Wa Yohana*, *Achikalata Jua Paolo Jua Ndumitume Kwa Wa Korinti* and *Kalata Jua Paolo Jua Ndumitume Kwa Wa Rumi* (Houston 2022: 9). There were also some primers, produced by R.S. Hynde in 1892 and 1894 (Houston 2022: 6).
5. The CLS dictionary is a monolingual dictionary produced as a tool for the hoped-for initiative of vernacular education in Ciyawo, which was made policy by the Malawian government in 1996 but never developed beyond a few pilot projects.
6. There is conjecture as to what constitutes a word, especially in Bantu languages, as it depends on the orthographical style (De Schryver and Prinsloo 2000: 100). What takes three words in English, "I am going", is written conjunctively as a single word in Ciyawo — Ngwawula.
7. Lorna Morris (née Hiles) was a tremendous help to our project as she had experience as an editor working on the *Oxford South African Illustrated School Dictionary* (2008).
8. I want to particularly thank my colleagues at Baptist Mission Australia, formerly Global Interaction, who saw the value of this project and supported it over many years, particularly Mrs. Alison Nissley and Dr. John Davis, who oversee projects in the organisation and who championed it and encouraged me along the way, as well as the many people who supported this project financially.
9. This figure was calculated from SFI giving records between 2009–2025.

References

- Abdallah, Y.B.** 1973. *The Yaos: Chiikala cha WaYao*. Edited and translated by M. Sanderson. London: Frank Cass.
- Agar, M.** 2002. *Language Shock: Understanding The Culture of Conversation*. New York: Harper Perennial.
- Atkins, S., J. Clear and N. Ostler.** 1991. *Corpus Design Criteria*.
<http://www.natcorp.ox.ac.uk/archive/vault/tgaw02.pdf> [16 May 2025]
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Baldauf, R.B and R.B. Kaplan.** 2004. *Language Planning and Policy in Africa: Botswana, Malawi, Mozambique and South Africa*. Vol. 1. Clevedon: Multilingual Matters.
- Banda, F., A. Mtenje, L. Miti, V. Chanda, G. Kamwendo, A. Ngunga, M. Liphola, C. Manuel, B. Siteo, S. Simango and M. Nkolola.** 2008. *A Unified Standard Orthography for South-Central African Languages: Malawi, Mozambique and Zambia*. Second edition. Monograph Series No. 229. Cape Town: CASAS.
- Barwell, K.** 1986. *Bible Translation: An Introductory Course in Translation Principles*. Third edition. Dallas: Summer Institute of Linguistics.
- Bayetto, A.** 2018. *The New Oxford Wordlist Research Report*.
https://www.oup.com.au/__data/assets/pdf_file/0016/120841/SCHL_WORDLIST_Report-2018_FA_2-web.pdf
- Centre for Language Studies (CLS).** 2005. *The Orthography of Ciyawo*. Chileka: E+V Publications.
- Centre for Language Studies (CLS).** 2013. *Mgopolela Malowé jwa Ciyawo: Ciyawo Dictionary*. Blantyre: Dzuka.
- Coxhead, A.** 2000. A New Academic Word List. *TESOL Quarterly* 34(2): 213-238.
- Dang, T.N.Y. and S. Webb.** 2016. Making an Essential World List for Beginners. Nation, I.S.P. (Ed.). 2016. *Making and Using Word Lists for Language Learning and Testing*: 153-167. Amsterdam: John Benjamins.
- Datereportal.** 2025. *Digital 2025: Malawi*.
<https://datereportal.com/reports/digital-2025-malawi> [17 May 2025]
- De Schryver, G.-M. and D.J. Prinsloo.** 2000. The Compilation of Electronic Corpora: With Special Reference to the African Languages. *Southern African Linguistics and Applied Language Studies* 18(1-4): 89-106.
- Dicks, I.D.** 2006. *Wisdom of the Yawo People: Under the Elephant's Belly You Can't Pass Twice*. Zomba: Kachere.
- Dicks, I.D.** 2018. *English-Ciyawo Learner's Dictionary*. Mzuzu: Mzuni Press.
- Dicks, I.D.** 2022. *Ciyawo-English Learner's Dictionary*. Mzuzu: Mzuni Press.
- Dicks, I.D.** 2024a. *Ciyawo-English Learner's Dictionary*. Smartphone Application. Mzuni Press / Google Play Store / Apple App Store.
- Dicks, I.D.** 2024b. *English-Ciyawo Learner's Dictionary*, Smartphone Application. Mzuni Press / Google Play Store / Apple App Store.
- Fetterman, D.M.** 2010. *Ethnography: Step-by-Step*. Third edition. Los Angeles: SAGE.
- Hornby, A.S. (Ed.).** 2010. *Oxford Advanced Learner's Dictionary*. Eighth edition. Oxford: Oxford University Press.

-
- Houston, T.J.** 2022. Utenga Wambone — The "Good News": An Exploration of Historical Ciyawo Bible Translations and Linguistic Texts. *Studia Historiae Ecclesiasticae* 48(3): 1-18.
- Kamwendo, G.H.** 2000. Interfacing Language Research with Policy: The Case of Language in Education in Malawi. *Nordic Journal of African Studies* 9(2): 1-10.
- Kandybowicz, J. and H. Torrence.** 2017. Africa's Endangered Languages: An Overview. Kandybowicz, J. and H. Torrence (Eds.). 2017. *Africa's Endangered Languages: Documentary and Theoretical Approaches*: 1-10. Oxford/New York: Oxford University Press.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. New York/Cambridge: Cambridge University Press.
- Macdonald, D.** 1881. *East African Tales*. Edinburgh: William Blackwood and Sons.
- Macdonald, D.** 1882. *Africana; or, The Heart of Heathen Africa*. Vol. 1. London: Simpkin Marshall & Co.
- Malawi Data Portal.** 2018.
<https://malawi.opendataforafrica.org/gesjee/population-2018> [17 May 2025]
- Matiki, A.J.** 2001. The Social Significance of English in Malawi. *World Englishes* 20(2): 201-218.
- Matiki, A.J.** 2006. Literacy, Ethnolinguistic Diversity and Transitional Bilingual Education in Malawi. *International Journal of Bilingual Education and Bilingualism* 9(2): 239-254.
- Morris, L.H.** 2021. *A Model for a Comprehensive Electronic School Dictionary for South African Primary School Learners*. Unpublished Ph.D. Dissertation. Stellenbosch: Stellenbosch University.
- Nation, I.S.P.** 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I.S.P.** 2022. *Learning Vocabulary in Another Language*. Third edition. Cambridge: Cambridge University Press.
- National Statistical Office Malawi.** 2019. *2018 Malawi Population and Housing Census. Main Report*.
<https://www.nsomalawi.mw/census/2018> [17 May 2025]
- National Statistical Office Malawi.** 2022. *The Second Malawi Multidimensional Poverty Index Report — Nov 2022*.
https://www.undp.org/sites/g/files/zskgke326/files/2023-08/Malawi-Multidimensional%20Poverty%20Index%20Report_0.pdf [17 May 2025]
- Oxford Learner's Dictionaries.** 2025. *The Oxford 3000 and the Oxford 5000*.
<https://www.oxfordlearnersdictionaries.com/about/wordlists/oxford3000-5000#> [17 May 2025]
- Prinsloo, D.J.** 2017. Africa's Response to the Corpus Revolution. Xu, Hai (Ed.). 2017. *Proceedings of the 11th International Conference of the Asian Association for Lexicography (ASIALEX 2017), 10-12 June 2017, Guangzhou, China: Lexicography in Asia: Challenges, Innovations and Prospects*: 20-31. Guangzhou, China: ASIALEX.
- Reilly, C.** 2019. Attitudes Towards English as a Medium of Instruction in Malawian Universities. *English Academy Review* 36(1): 32-45.
- Shaffer, D.E.** 2005. Review of *Oxford Advanced Learner's Dictionary of Current English* (7th ed.). Korea TESOL Journal 8(1): 135-141.
- Sinclair, J.** 2004. Corpus and Text: Basic Principles. Wynne M. (Ed.). 2004. *Developing Linguistic Corpora: A Guide to Good Practice*: 1-16. Oxford: Oxbow Books.
<https://users.ox.ac.uk/~martinw/dlc/> [16 May 2025]
- Trudell, B.** 2016. *The Impact of Language Policy and Practice on Children's Learning, Evidence from Eastern and Southern Africa*. Nairobi: UNICEF.
<https://www.unicef.org/esa/sites/unicef.org/esa/files/2018-09/UNICEF-2016-Language-and-Learning-FullReport.pdf> [16 May 2025]

Trudgen, R. 2000. *Why Warriors Lie Down and Die*. Darwin: Aboriginal Resource and Development Services Inc.

UNESCO. 2024. *Malawi: Education Country Brief*.

<https://www.iicba.unesco.org/en/malawi> [17 May 2025]

Webb, S. and P. Nation. 2017. *How Vocabulary is Learned*. Oxford: Oxford University Press.