# LEXICOSTATISTICAL ANALYSIS AND YAMBA ORTHOGRAPHY DEVELOPMENT

**By**

**Talah Benard S, Vincent A. Tanda and Wega A. Simeu**

**Abstract**

*This study sets out to examine the importance of lexicostatistical analysis in the development of an acceptable orthography for multi-dialectal communities like Yamba with 17 dialects. Many intractable orthography problems prevail with the extant Yamba standard reference orthography. These problems stem from the acceptability factor to the usage of materials produced in the language using the Mbem dialect as the reference dialect. The data for this work was collected using an eclectic method. To answer the main research question of this paper which is: how can lexicostatistical analysis aid in orthography development in a multidialectal languages, questionnaires, structured interview, Swadesh wordlist and archive were used as the research instruments. The objectives of this paper are to examine the validity of lexicostatistical analysis in orthography development for multidialectal languages and to propose an alternative way of developing orthographies for multidialectal languages. The analysis of the data was done using Transformational Generative Grammar and Generative phonology theories. The results indicated diverse intelligibility levels between all the dialects of Yamba. Developing supplementary orthographies for dialects with as low as 30% intelligibility supports the use of the orthographies to do mother tongue literacy. Therefore, orthography development for languages with multiplicity of dialect can be achieved only after a thorough lexicostatistical analysis that establishes the levels of intelligibility between the dialects.*

Keywords: *lexicostatistics, orthography, generative grammar, literacy, mother tongue*.

---

## 1. INTRODUCTION

Research on orthography development in recent years has been characterized by the use of reference dialects for standardization (Simons and Fennig, 1990). They postulate that dialect variation may be as great in a language as to prevent speakers of different dialects of the same language from understanding one another. It is worthy of note that the goal of language development is to promote literacy and mother tongue education and the perpetuation of the use of the language in question universally. But, sometimes the writing system put in place (especially in multi-dialectal situation) is either under-used or not used at all. This occurs with different writing systems throughout the world and especially in Africa, considering that one dialect in multi-dialectal languages gets developed as the standard version. The standardization efforts met with frustration, sometimes by the fact that a majority of speakers of such languages become reticent towards using the standardized version in reading and writing the language. Lexicostatistical analysis of Yamba will x-ray the divergent intelligibility levels between the seventeen presumed dialects which seems to be at the root of low literacy rate in Yamba after more than three decades of standardisation (Bradley, 1986)

Many languages are endowed with different variants, which necessitates the selection of one variant for standardization. For the case of Yamba, Mbem dialect was chosen and developed as the standard or reference dialect following theories and criteria laid down by (Bos et al, 2004:4), (Sadembouo, 1980) sanctioning the selection of a reference dialect. The present orthography or alphabet of Yamba have been used to promote literacy for over thirty years and a number of difficulties have been observed. Though much literacy efforts were made at the inception of Yamba orthography development, the fortunes dwindled with time as the question of inter-comprehension gradually surfaced. Simons and Fennig, (2018) states that about 10% of the population is literate in the mother tongue out of a total population of about 80,000 people.

The fact that the didactic materials and the New Testament of the Bible are all produced in the reference dialect causes enormous problems to speakers of the 16 other dialects. The problem is further compounded by the fact that most speakers of the 16 dialects have a limited comprehension of the reference dialect as Bradley, (1986) states:

At the beginning of linguistic work on the language, people immediately began to      inform     me that each village had a different dialect of Yamba, some very different from      the Mbem one. From these comments I received the general impression that some of      the     furthest-out     villages     from     Mbem, geographically speaking, were not mutually   intelligible with the Mbem dialect"

Despite these complaints, the Mbem dialect was standardized as the reference for the Yamba speakers. A lexicostatistical analysis would have shade more light on the inter-comprehension and intra-comprehension

levels of Mbem and the other sixteen dialects. ). This paper, lexicostatistical analysis and Yamba orthography development seeks propose sustainable solutions to this problem by examining the validity of lexicostatistical analysis in orthography development for multidialectal languages and propose an alternative way of developing orthographies for multidialectal languages. Reviewing some literature on orthography development lays the basis for better comprehension of the lapses of orthography development without a proper lexicostatistical analysis.

## 2. REVIEW OF SOME WORKS

This focuses on the works of lanauge development generally, with specific attention on Yamba. According to Easton (2002)  "the ownership of the orthography…is in the hands of the speakers of a language. They must make the decisions regarding how the orthography should look and how to spell the language". He posits that instead of delivering an orthography to a language community, it is better to involve them in the process so that everyone can accept the decisions made. When a community requests for help in their language development effort, the best way to help them is through workshops. In addition, since the people do not see the orthography as their own, they do not identify with the language and therefore do very little to carry out its development. He fails to recognize the need for an orthography that can help languages in cluster or dialects of the same language with low intelligibility level with the reference dialect. It is therefore incumbent to do a lexicostatistical analysis in order to find a solution that is sustainable and workable.

Bradley significantly worked on the language since she has remained the main linguist dedicated to the development of the Yamba language. She discusses serial verbs and clauses in Yamba (Bradley, 1982). She uses only the Mbem dialect to demonstrate serial verbs in the lanuage. In her write up she notes that the language lines up verbs in the same clause but that they do not enter into coordinate or subordinate relationship with each other. This work uses Mbem as the reference dialect. What is not examined is whether verbs behave in like manner in the rest of the dialects. Therefore the need for a comparative study of Mbem and other dialects.

In addition,  the alphabet and orthography statement Yamba was elaborated, (Bradley, 1986a), in which the writing rules of the Yamba language are spelt out. This work lays the foundation for literacy and multilingual education in Yamba.  The alphabet of the language, the orthographic rules, morphophonemic rules and how tones are written in the language are elaborately spelt out. Salient is also the fact that she basically uses only the Mbem dialect with little or no inclusion of the other sixteen (16) dialects in the writing system.

To further develop materials for mother tongue literacy and multilingual education, Bradley (1986b) produced the "the Yamba reading and writing book". Considering the little demonstrated in the use of this material, she decided to investigate the effects of the mulitple dialect situation on the usage of the primer (Bradley, 1986c). This investigation was in attempt to either confirm or refute earlier discovery. This affirmation of lack of intelligibility amongst the dialects is indicative of the reasons for the non usage of materials produced using the Mbem dialect by speakers of the other dialects. This puts to question the selection process of Mbem without sufficient data to prove its general comprehension.

Also, Nzenge (2001) did a morphosyntatic study of the basic sentence structure in Yamba. Just like Bradley, Nzenge uses Mbem as the reference dialect and posits that all materials should be produced using it. Her focus was on the morphosyntatic rules of the language, though using only Mbem without taking into consideration the possible differences that exist between the reference dialect and a majority of the other dialects. Interestingly, she states that there are 10 dialects in Yamba instead of 17 earlier mentioned by Bradley (1982, 1986a) and (Terry, 1981). It is therefore evident that her work did not cover the entire area where the language is spoken thereby neglecting what happens with the other 7 dialects. In additition, she doesn't make mention of the intelligibility level between the dialects. In order to lay the basis for proper analysis, some basic grammar and phonology theories were used to examine the intelligibility levels.

## 3. Theoretical framework

Two main theories sanctioned this paper. Transformational generative grammar and generative phonology. Transformational grammar is a system of language analysis. It is part of generative grammar, with focus on natural languages. It considers grammar to be a system of rules that generate grammatical sentences in particular languages. Transformational Generative Grammar theory has two main working core concepts; the innate linguistic or language knowledge and grammaticality.
Generative phonology theory is closely related to transformational grammar theory.
Generative phonology aims at establishing a set of rules, principles, or constraints that are capable to produce surface phonetic forms of a language. It also models the internalized linguistic or language knowledge of native speakers (Chapman and Routlege, 2009:77). Lexicostatistics makes the relationship between language rules and grammar explicit. It is on this basis that language rules are made. The data collection method for this paper also needed to be spelt out.

## 4. Methodology

The methods of data collection included participant observation, semi-structured interviews and Swadesh wordlist to collect words and phrases and transcribed using the International Phonetic Alphabet (IPA). These words were further analysed phonologically using Chomsky (1965)'s generative phonological theory, generative grammar theory to identify phonemes of the dialects which are compared with those of the reference dialect (Mbem) and the intelligibility levels established. The analysis of this data resulted in a number of findings are presented below.

## 5. Results and findings

The analysis produced divergent intelligibility results for Yamba, which influences the way forward for the standardization of Yamba language.

The table below indicates the lexical distance between Mbem and the other dialects.

Table 1: comparative analysis of roots of Yamba words

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **English** | **Mbem** | **Gamfe** | **Ngung** | **Ntong** | **Gwembe** | **Bom** | **Sih** | **Yang** | **Rom** | **Nkot** | **Nwa** | **Gom** | **Kwak** | **Ntim** |
| all | kpép | pwə́' | Lăŋ | pwét | píə́ | lŏŋ | lŏŋ | lŏŋ | mɨ́ | lîŋ | pwé' | pwə' | lŏŋ | lŏŋ |
| and | nə́ | ábə́də́ | bā | bə́ | bə́ | ŋgani | bɔ́n | bə́ | | ndə́ | bə́ | bə́ | bɔ́n | mbĕn |
| animal | nàm | nàm | nyē | Nyàm | nàm | nàm | nàm | nàm | nyàm | nyàm | Nàm | nyàm | nàm | nĕmŋgɔ̀n |
| ashes | vɨ́ | víə́ | dùdùh | vɨ́ | víə́ | vɨ́ | víə́ | vɨ́ | vīə | vī | vɨ́ | | víə̄ | mvú |
| at | fə́/mə́ | jyá | á | | kó | | wán | ndzə́ | | tə́ | sə́/kó | | ndə́ | |
| back | ŋkwèn | máŋkwè | kɛ̀ | Nkwàn | ŋkwè | ŋkɔ̀n | ŋkɔ̀n | ŋkòn | kwān | fɔ́kɔ́ | ŋkwèn | ŋkwàn | kɔ̀m | ŋkɔ̀r |
| bad | bíp | bép | kĕ | mátə̂m | bíp | bíp | bíp | bíp | | bíp | bíp | bīp | bíp | táhá |
| because | bǐ | ŋgə́bɨ́ | ndɨ́ | Jó | bǐjā | njó | njó | njó | bú | | njónə̀ | bɨ́ | njó | njó |
| belly | və̀m | vɔ̀m | və̀h | və̀m | və̀m | və̀m | və̀m | və̀m | və̀m | və̀m | vùm | və̀m | və̀m | mvùm |
| big | gháŋ | ŋkɨ́ŋ | méigbō | nàsə̀bə̀t | kɨ̌ | máháŋ | yə́r | yə́r | nə̄yám | mágháŋ | nə̀gháŋ | nə̄mēn | rɨ̌n | nə̀mɨk |
| bird | swíŋ | síŋ | síŋ | Shíŋ | swíŋ | síŋ | síŋ | síŋ | síŋ | shíŋ | swíŋ | shíŋ | síŋ | síŋ |
| bite | lúm | lúm | lə́m | Lúm | lúm | lúm | lúm | lím | chíchík | lúm | lúm | lúm | lúm | līp |
| black | bìp | fɨ́fɨ́ | méibì | másə́n | fɨ́fɨ́ | sə́sə̂n | másə́n | sə́sə́n | sə́sə̂n | sə́sə̂n | sə́sén | sə́sə̂n | sə̄sə́n | măbíp |
| blow | fɨ́ə́p | ndò | fíə̄h | fíə́p | fə́p | fíə́p | fíə́p | fə́p | fíə́p | | fíə́p | fíép | fíə́p | fɔ̀p |
| bone | və́p | və́p | və̄h | və́p | mbɨ́ | və́p | və́p | və́p | və́p | və́p | vóp | və̄p | və́p | |
| breath | yɔ̀ŋsə̄ | yɔ̀ŋsə̀ | lūŋ | máyɔ̀ŋsə̀n | yɔ̀sə́ | yɔ̀ŋsə̀ | yɔ̀ŋsə̀ | yɔ̀ŋsə̄ | yɔ̀ŋsə́ | yɔ̀ŋsə́ | yɔ̀ŋsə́ | yɔ̄ŋsə́ | yɔ̀ŋsə̀ | |
| few | jó | jwó | ntwíh | Móntwírt | jó | ndwírt | mŏntə́k | jó | jó | ndwít | jó | ndwít | mɔ̂ntə́k | njét |
| fight | lǐ | lì | mbwó | lǐ | lǐ | Mbó | mbó | lŭ | lī | lǐ | lǐ | mbó | lŭ | lū |
| fire | mís | mís | méh | Mít | mís | mə́s | mə́s | mə́s | mís | mbɨ́ə́s | mís | méh | mís | mát |
| fish | fúk | fúk | sīe | Fúk | fúk | Fúk | fúk | fúk | fúk | ŋgŏŋ | fúk | sē | fúp | zú' |
| five | tàŋ | tàŋ | tôŋ | Tán | tàŋ | tân | tân | tàŋ | tân | tân | tàŋ | tŏŋ | tân | twŏŋ |
| flow | cháŋ | kì | kǐ | Cháŋ | | Cháŋ | və̀rə́ŋ | chén | sɨ́ | chyá | chyá | cháŋ | cháŋ | |
| fly | ndzə̀ŋ | nchi | líhɨ́ | lí'kə́ | lí | bwɨ̌m | líhə́ | dzɔ̀ŋə́ | líkə́ | líɛ'kə́ | dzɔ̀ŋə́ | léhé | bwɨ̌m | njì |
| neck | mì | mī | mīh | mī | mì | mì | mī | mì | mī | mī | mì | mī | mī | mítɔ́n |
| new | fi | nə̀fi | nə̄fi | Fí | fi | fi | nə̄fi | fi | fi | fi | nə̀fi | nə̄fi | nə̀fi | nə́fi |
| night | ndz | njə́ | ndzə́h | ndzə́m | njím | mád | ndz | ndzə́ | ndzə́ | ndə́m | ndzə́ | ndzə́ | ndə́m | míndí |

| | ə́m | m | | | | zə́m | ə́m | m | m | | m | m | p |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| nose | ndòzé | wí | zē | ŋtàsē | wí | ntèzə́ | tāsə́ | ndɔ̀zə́ | ndɔ̀zə́ | zē | ndìèzə́ | ndòzē | tàsə́ | ngɔyú |
| not | kà | kábá | kə́ | Ká | kà | kàwâ | kà | háí | háí | kā | ká' | kān | ká | hàí |
| old | jə̀wìs | rìέ | ŋgέ | mágíŋ | rìə́ | rěŋ | ràŋ | rìrǐŋ | rēn | rěn | rěŋ | rěn | márέŋ | mbóp |
| one | mɔ̀'fís | mɔ̀'fis | mōhféh | mōtfìt | mɔ̀fis | mɔ̀'fís | mɔ̀hfέs | mònfέs | mō'fís | mɔ̀'fís | mɔ̀'fís | mō'feh | mɔ̀hfís | mɔ̌rt |
| other | mɔ́k | mɔ̌wì | mɔ́h | sə̀mɔ́k | mɔ̀háné | nὲmɔ́k | mɔ́k | mɔ́k | mɔ́k | mū'ú | wó | mɔ́h | mɔ́k | būmlɔ̌ŋ |
| person | ŋgὲngàŋ | ŋwὲ | ŋwēn | ŋwēn | bǐyā | ŋwὲmɔ́k | ŋwὲn | ŋwὲn | ŋwὲn | ŋwūn | ŋwù | ŋwēn | ŋwὲn | ŋwìn |

Examining mutual intelligibility levels between the dialects of Yamba is necessary. A lexicostatistical analysis revealed different levels of intelligibility between Mbem and the rest of the 16 dialects. The table above shades light on the intelligibility levels of the reference dialect with the other sixteen (16) dialects. The lexical items used to get these results were taken from a comparative analysis of Swadesh (1955)'s word list. From the roots of the words, varying degree of differences can be noticed. Though the dialects share a lot of lexical similarity, a closer look reveal differences in the sounds or pronunciation. In addition, the lexical distance between the words equally shaded light on the intelligibility index of the dialects.

Table 2: lexical distance between Mbem and the other dialects

From table above, the claim made by some respondent that not every one understands Mbem is

| Dialect | Mbem | Nwa | Rom | Gom | Bom | Nkot | Gwembe | Gamfe | Sih | Ntong | Ngung | Yang | Ntim |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mbem | | | | | | | | | | | | | |
| Nwa | 75 | | | | | | | | | | | | |
| Rom | 79 | 51 | | | | | | | | | | | |
| Gom | 75 | 58 | 65 | | | | | | | | | | |
| Bom | 74 | 56 | 38 | 61 | | | | | | | | | |
| Nkot | 71 | 49 | 60 | 56 | 58 | | | | | | | | |
| Gwembe | 76 | 57 | 53 | 61 | 54 | 47 | | | | | | | |
| Gamfe | 67 | 58 | 54 | 51 | 48 | 53 | 77 | - | | | | | |
| Sih | 67 | 67 | 64 | 58 | 67 | 59 | 49 | 51 | - | | | | |
| Ntong | 62 | 64 | 56 | 57 | 52 | 58 | 50 | 47 | 72 | - | | | |
| Ngung | 57 | 49 | 53 | 52 | 51 | 53 | 37 | 38 | 64 | 63 | - | | |
| Yang | 60 | 73 | 58 | 49 | 65 | 57 | 54 | 61 | 57 | 51 | 49 | - | |
| Ntim | 35 | 39 | 36 | 21 | 34 | 31 | 30 | 24 | 21 | 30 | 21 | 29 | - |
| Kwak | 72 | 48 | 57 | 24 | 57 | 46 | 48 | 44 | 65 | 57 | 48 | 51 | 30 |

evident. The dialects at the borders with other language show varying degree of intelligibility with Mbem. Though the items used for analysis were mostly lexical entries, some conclusions about the mutual intelligibility of Mbem and the rest of the dialects can be made. The dialects at the peripheries are Ngung with a mutual intelligibility percentage of 57%, Ntim, 35%, and Nkot, 71.4%. Bradley (1986b) had earlier claimed that these dialects were influenced by neighbouring languages like Mfumte, for Ngung, Limbum for Nkot and Mambilla for Ntim (and Saam whose data is not available) respectively. One interesting discovery was the very low percentage of Ntim, (neighbour to Saam) which Bradley (1986b:5) indicated that the speakers of these dialect will find it very difficult to use materials produced in Mbem. An examination of the lexicostatistical analysis suggested that Ntim (and probably Saam) are surely different languages that borrowed heavily from Yamba.

The nucleus of the lexical items had very few similarities with other dialects. However, the high percentage of lexical differences between Mbem and Ntim is already a pointer to the fact that Ntim and Saam maybe different languages. Despite this conclusion about boarder dialects, Nkot still indicated a high intelligibility percentage with Mbem (71.4%). This means that, Nkot speakers will relatively understand Mbem and can use material produced in Mbem for literacy and mother tongue education. Their reticence towards the orthography in Mbem, then it is a sociolinguistic problem. More tests needs to be carried out with Nkot and Limbum to ascertain the influence of Limbum on the dialect.

The second group of dialects with varying degree of intelligibility are Gom with 75%, Gwembe, 78.7, Bom 74, Rom, 79%, Kwak, 72.5, and Nwa, 75%. These dialects are closer to Mbem and the high mutual intelligibility indexes. Notice that these five dialects have a more than 70% lexical similirities with Rom standing out. The next group of dialects are Gamfe, 67%, Ntong 62.5% Sih, 67%, Yang, 60% that are further away from Mbem with great amount of variation in pronunciation and vocabulary. Unfortunately, there was no data for Saam, Mfe and Rom, which is part of this group. This group has some differences in phonology and grammar, but can easily understand the Mbem dialect given that mutual intelligibility of this group and Mbem is the highest. These dialects are the clcosest neighbours to Mbem. More efforts needs to be deployed to promote literacy and mother tongue education in these villages. Despite the fact that mother tongue literacy classes had been organized to promote reading and writing in the language, the literacy rate has not exceeded 10%. The low mutual intelligibility level with some of the dialect has surely contributed to this.

The figure below summarises the intelligibility index between Mbem and the other thirteen (13) dialects.
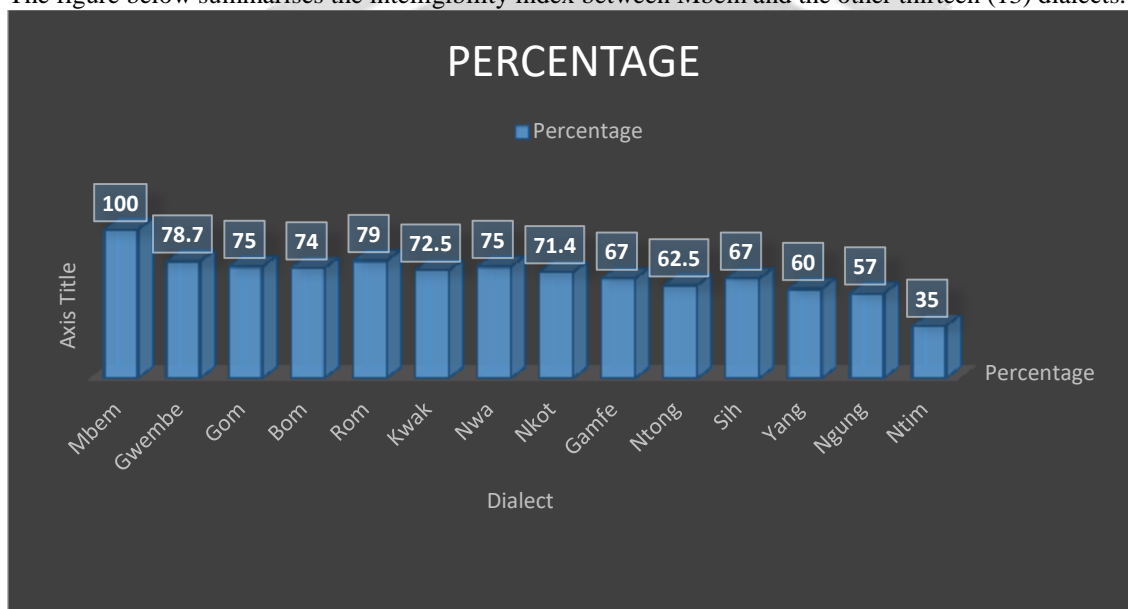


Fig. 1: lexical distance between Mbem and other dialects.

Seguin (2004), summarises intelligibility index with varying degrees and its implication on the choice of a reference dialect. Table 2 below presents it as follows:

Table 3: Intelligibility index for reference dialect

|  | Percentages of similar words | |
|---|---|---|
| **Inherent intelligibility** | **Above 60%** | **Below 60%** |
| Above 80% | Very similar speech varieties maybe referred to as similar dialeccts if inherent intelligibility is high | Several dissimilar or slightly similar speech varieties maybe transferred to different languages (no dialect intelligibility testing is required) |

| Below 80% | Several very similar speech varieties maybe referred to either as dissimilar dialects or different dialects if inherent intelligility is low | |
|---|---|---|

Source: Seguin, 2004.

From table 2 above, a mutual intelligibility of eighty percent (80%) and above between dialects, is considered inherent intelligibility. This means that, the dialects are highly mutually intelligible and are therefore same language. When the percentage is below eighty percent (80%), it means that, there are several very similar speech varieties, therefore considered as different dialect with very high intelligibiligy. However, if the percentage is slightly above sixty (60%+) or 70%, it is an indication that there are several very similar speech varieties that can be reffered to as different dialects of thte language. Therefore, the intelligibility is low and way to promote literacy and mother tongue education needs to be put in place. This is the premise on which we are proposing supplementary orthographies in such cases.

On the other hand, if the intelligibility between the dialects is below sixty percent (60%-), there is many dissimilar speech varieties or low similar speech varietieis, and therefore, there are different languages. Taking this into cocnsideration, the intelligibility between the 16 dialects and Mbem records varying degrees of intelligibility. Meanwhile, Ngung, Ntim and probably Saam record intelligibility of less than sixty percent (60%) with Mbem, a majority of the dialects, (Ntong, 62.5%, kwak, 72.5%, Yang, 60%, Gom, 75%, Gamfe, 67%, Gwembe, 78%, Nkot, 71.4%, Sih, 67%, and Bom, 74, Rom 75% ) fall below the 80% magin suggested by Seguin. The only dialect that meets the criteria of above 80% is Nwa with an intelligibility index of eithgty percent (80%). These percentages are telling of the low mother tongue literacy rate in Yamba over the decades.

Similarly, Wega, 2004, suggests that if mutual intelligibility level exceeds fifty (50%) they are dialects of the same language. If that suggestion is taken into consideration, only Ngung (57%) and Ntim (35%) and probably Saam will meet this condition of being different languages. Since one of our objectives is on how to make an orthography acceptable (research question 3), the development of supplementary orthography will be the proposed solution (research questions 4).

Salient to note here is also the level of mutual intelligibility for dialects to be considered inherent suggested by Simons and Fennig (2021). To them, the mutual intelligibility must be between 80% and 85% between dialects in order for them to be considered as same language.

Apart from establishing the mutual intelligibility between Mbem and the other dialects, an attempt was made to establish the inter-comprehension between the other dialects.
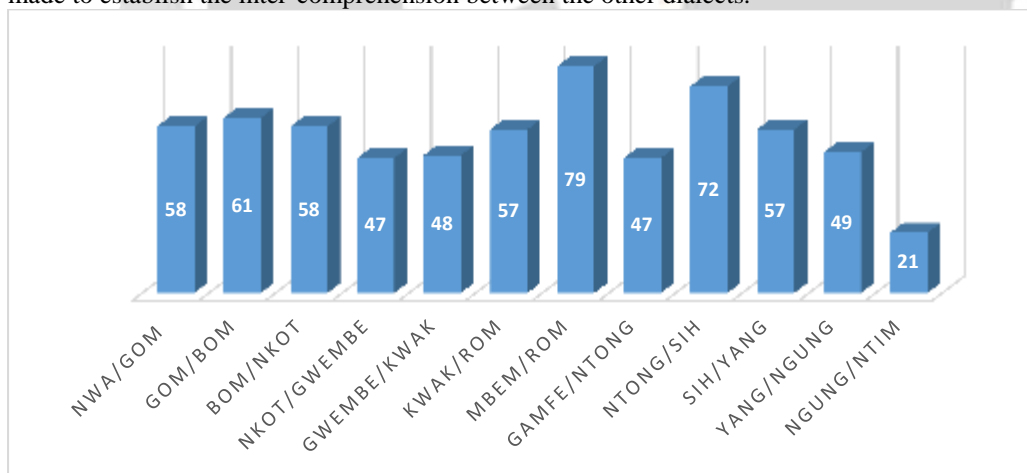


Fig. Mutual intelligibility of other Yamba dialects

The lexicostatistical analysis revealed a number of issues. The mutual intelligibility index between the dialects as seen in the figure above is as follows:

1.  The Nwa dialect: apart from having a high intelligibility with Mem (75%), also has a relatively high intelligibility with Yang, 73%, Sih, 67%, and Ntong, 64%. Its intercomprehension with Gom, Bom, Gwembe, Gamfe, and Rom (58%, 56%, 57%, 58% and 51% respectively) is lower than the first three. The third group consists of dialects that are further away from Nwa. There are Ngung, 49%, Kwak, 48%, Nkot, 49% and Ntim, 39%. Note, that Nwa is the sub divisional headquarter but it's influence is limited or reduced due to the cosmopolitan nature of Nwa.
2.  Rom dalect: it equally shares diverse intercomprehension fortunes with the rest of the dialects. Apart from Bom, 38%, and Ntim, 36% lexical similarity, the intelligibility index of Rom and the rest ranges between

50% and 60%. The fact that Rom shares meaningful boundaries only wth Mbem and Gom. A majority of the dialects are further away from it.

3. The Gom dialect shares a high intelligibility with the reference dialect (75%). It was however, revealed that it shares a relatively low intercomprehension with the rest of the dialects. Gom enjoys intercomprehension of 60%+ only with two dialects (Bom, 61% and Gwembe, 61%). It shares between 50% and 58% with Ngung, Nwa, Ntong, Rom, Nkot, Gamfe, Sih, and Yang. Three dialects (Ntim, Kwak, and Yang) present a similarity index of less than 50%. This is explained by the fact that those dialects are further away from Gom.

4. The Bom dialect shares between 50% and 65% with Rom, Nwa, Gwembe, Sih, Ntong, Ngung, Yang, Nkot, Gom, and Kwak, while sharing a lower than 50% similarity index with Ntim, and Gamfe.

5. Gamfe and Gwembe share a lot. Both of them share a lexical similarity of 77% meanwhile it dwindles as one moves to the dialects that are far from them. The Kwak, Ntong, Bom, Ntim, Ntong, and Ngung present an interocmprehension index of less than 50%. The rest of the dialect reveal the same scenario as they with Bom.

6. Yang, Nkot, and Kwak demonstrate similar intelligibility index with the rest of the dialect but for the fact that Nkot and Yang are at the boarders of Yamba sharing boundaries with Limbum and Mambilla respectively. While Nkot shows a lower lexical similarities with Nwa, 49%, Gwembe, 47%, Kwak, 48%, and Ntim 30%,, Kwak does same with Gamfe, 44%, Ntim, 30%, Nwa, 48%, and Gwembe 48%.

It is important to note that the above dialects are those Bradley (1986b) classified as a group of eleven (11) clustered around the Mbem dialect. She claimed that their speakers could use materials produced in Mbem dialect with relative ease.

       The second group of dialects (Yamba North dialects), are those considered to be further away from Mbem. Bradley (1986b) posited that their speakers would have difficulties using the Mbem orthography in literacy and mother tongue education. They all present diverse lexical similarity indexes with one another.

1. Ntong's similarity index vary from 30% with Ntim to 64% with Nwa. Its lexical similarity are generally low with two dialects, which includes Ntim with 30%, Gamfe, 47%. The index shows two dialects with percentages slightly above 60%. These are Nwa with 64%, and Ngung with 63 % similarity. Interestingly, Ntong has the lowest percentage with Yamba North dialects. The rest of the dialects fall within 50% and 58% range of lexical similarity.

2. The Sih dialect enjoys a higher intelligibility with the Yamba North group of dialects. While it shares a lexical similarity of 72% with Ntong, which is a Yamba North dialect, it shares a relatively comfortable intelligibility with Ngung, 64%, Kwak, 65%, and Bom, 67%. It shares lexical similarity of 51% with Gamfe, 57% with Yang, 58% with Nwa and 59% with Nkot. Sih is the dialect with the highest intelligibility index among the group of six dialects of Yamba North.

3. Similarly, Ngung shows diverse intelligibility indexes with the rest of the dialects. Ngungs shares an above 50% intelligibility index with Nkot, Bom, Ntong and Sih, it shares a lower percentage with four dialects (Gamfe, 38%, Yang, 349%, Nwa 49%, Kwak, 48%, Gwembe 37%, and Ntim 21%. Out of these dialects, Gamfe, Yang, Gwembe, Kwak, Gamfe and Nwa are dialects clustered around Mbem.

4. The Ntim dialect looks more of a stranger in Yamba because its similarity indexes are all time low with almost all the dialects. Its highest percentage is 39% similarity with Nwa. It is evident that using the Mbem dialect materials for literacy and mother education would yield very little. Ntim is separate language and not a dialect of Yamba.

The analysis reveal that Sih has the highest intelligibility with dialects that are clustered around Mbem while Ntim illustrates the lowest of them all. It is therefore necessary to further examined Ntim's (and by extension Saam's) classification as a Yamba dialect. The diverse intelligibility indexes for forteen (14) of the 17 dialects justifies research question two (2). The reasons for the low literacy in Yamba after more than three decades seems to settle with dialectal mutual intelligibility besides other supporting factors.

       Taking into consideration past groupings by Bradley (1986a), the dialects that are close to Mbem can easily use the Mbem dialect as the standard reference dialect. In this case, seven (7) dialeccts namely; Nwa, Gom, Bom, Nkot, Kwak, Rom, and Gwembe can be regrouped with Mbem. This illustrated in the dialect hierarchical index below.
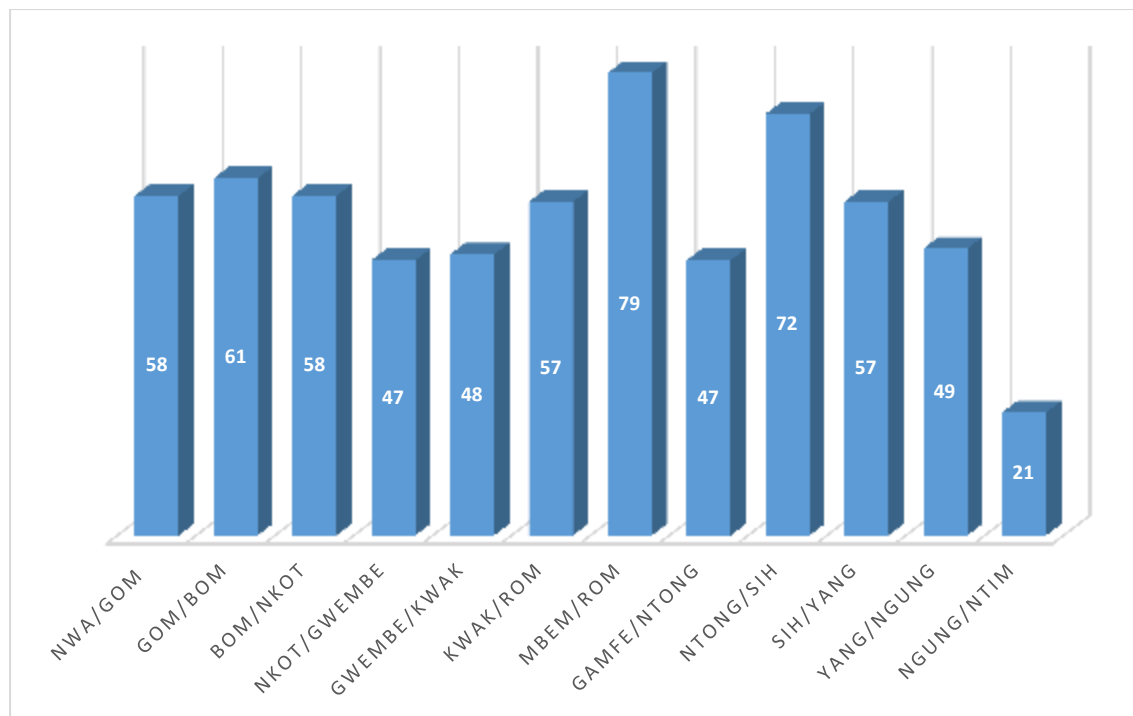
Fig: Hierachical index of dialects

It was observed that intra-comprehension between the dialects vary from one case to the other. Mbem and Rom revealed very high intelligibility with a percentage of 79%, meanwhile Ntim and Ngung have the lowest inter-comprehension. It also revealed that Sih and Ntong are very close with a 72% intra-comprehension. If we were to go by the suggestion made by Seguin (2004), the disparity in the intra and inter-comprehension is indicative of challenges of using one dialect to promote literacy and mother tongue education in Yamba. Dialects that demonstrate a seventy percent (70%) intelligibility can easily be put together with the reference dialect. This led to the recommendations below.

**6. Recommendations**

Taking into consideration the issues revealed by the lexicostatistical analysis, a number recommendations were made, which include but not limited to;

**6.1. Community based orthography development**

One of the criteria for the selection of a reference dialect is the unanimous acceptance of one variant by all the speakers of the language (Sadembouo, 1980). This dynamics is informed by the calibre of people who made the choice. This remains an ideal scenario which is an uphill task for the linguist to obtain. Since the choice of a reference dialect in Yamba has not obtained the desired results, examining the dynamics of linguistic awareness within the difference dialects of Yamba was necessary.

Speakers of many of the dialects are beginning to develop their own writing systems inspite of the advanced language development work done in Mbem. Taking advantage of this effort is primordial. When speakers of a language or dialect community is making efforts in language development and are ready to develop their language or dialect, getting their involvement is easier. Bos et al. (2010) outlines, among others, the holding of language awareness meetings, the organization of orthography development workshops, the formation of the language committee, the testing of the orthography and formal approval in the Kuy orthography development process in getting the community involved. Working with three dialects at the same time during data collection of this research proved very fruitful and beneficial to both the community and the research. The use of SIL African comparative word list facilitated the collection of multiple data within record time. Since acceptability of an orthography is salient in language development, those meetings in workshop made people to see the importance of participating in developing their language in any form.

In order to achieve acceptability, involving the language community becomes a condition sine-qua non. This involvement is not limited to physical participation but to financial and material involvement. During data collection for this research, many people requested money, claiming their community was already doing something for itself. But, when they realized the team put up by the working was working with us, the process was made easy. Worthy of note is that language communities trust their own consultants more than the researcher or linguist developing their language. The present orthography did not involve the entire community given that they felt alienated when Mbem was the only dialect retained. Seemingly little efforts were made to

get them involved directly, reason why their resorted to developing their dialects. The entails the use of local resources.

## 6.2. The use of local resources in orthography development

Every linguistic community is endowed with resources that can be very useful in developing an orthography for such a language. Apart from material and financial resources, many communities have human resources that can go a long way to support language development. In this, a community can decide to train one of theirs to develop the orthography of the language. Many of such cases were found in the Ntong, Ngung, and by extention, the Gom communities. The intriguing part of the scenario is that the training they had was insufficient to do a good linguistic work.

Bos et al. (2010) proposed the holding of language awareness meetings in the process of selecting a reference standard dialect for a multidialectal language. These awareness meetings culminates in the creation of a language committee that pilots the process of language development in the community and beyong. This committee can easily mobilze resources needed to develop the language. The community takes full responsibility for orthography development and doing basic didactic materials which is used for literacy. External help comes as an aide to their effort.

Another local committee that can be used to raise resources for language work is the church. Language development agencies in Cameroon like Summer Institute of Linguistics (SIL), The Cameroon Association for Bible Translation and Literacy (CABTAL), the Bible Society, World Team, just to name but these, work very closely with the churches in every community where they facilitate language development. The formation of the interchurch committee is prefaces their work since their focus is usually the translation of the Bible. Missionaries who brought the Gospel to Africa generally and to Cameroon in particular initiated this. These missionary organizations have continued to use translation of the Bible into local languages as a strategy to propagate the message of the Gospel. The interchurch committee is made of heads or their representative of every church denomination found the said community. They can raise resources from the various churches to support language development. Involving them in the entire process of orthography development and its subsequent ensueing didactic materials makes the difference. Therefore, synergizing with the language committee (language academy) and the Interchurch committee facilitates acceptability of the writing system. This gives the linguist ample time to play a coordinating and facilitating role with the local linguists who are MT speakers. This facilitates the promotion of supplementary orthography development efforts.

## 6.3. The promotion of supplementary orthography development efforts

This section deals with the analysis of individual community effort in language development. The quest and the desire for the codification of every language is at its prime as many communities are attempting to develop writing systems with little or no external help. Some of these efforts come because of either not being able to use the present writing system put in place using the standard reference dialect model or purely because of pride for their dialect. Examining the acceptability question revealed a number of interesting effforts going on with the speakers of Ntong, Ngung and Gom. They started developing their own writing systems and proceeded with a yearly publication of the annual diaries (which are all in circulation and use). Instead of discouraging these efforts, building on them and directing positively is an ideal decision to take.

The promotion of supplementary orthographies in dialects that have the human, financial and material resources to develop sub-writing systems will build confidence, inclusiveness and a sense of belonging. If the language committee guides these efforts, it will lead the speakers of the said dialects towards the use of documents produced in the reference dialect.

## 7. Conclusion

Considering the results obtained through this research, it is clear that lexicostatistical analysis of multidialectal languages is strongly recommended before any orthography development is done. Lexicostatistical analysis nip some problems in the bud and prevent resource wastage. It helps to establish a rapport between the researcher (linguist) and the speech community, thereby promoting participatory language development. This model is more sustainable that the "linguistically right" way of choosing the reference dialect mostly on sociolinguistic factors.

**References**

Bikoi (2012) Atlas Linguistique du Cameroun. CERDOTOLA: Yaounde, Cameroon.

Bos et al. (2010). Community-based orthography development: experiences from the    Kuy.    International Cooperation for Cambodia.

Bradley, V. (1982). Serial verbs and clauses in Yamba. Yaounde: SIL International.

…..        (1986). Alphabet and orthography statement of Yamba. Yaounde: SIL, Cameroon.

…         (1986). A survey of the Yamba dialect situation and its effects on the use of the primer. Yaounde: SIL
            Cameroon.

….         (2007). Yamba orthography guide. SIL, Yaounde, Cameroon.

Chomsky, N. (1956). Three models for the description of language. *Ire transaction on      information theory
2(3): 113-124, New York.*

---        (1965). Aspects of the theory of syntax. *Cambridge Mass, Journal of      Linguistics , volume 3 , issue
1 , april 1967, pp. 119 - 152*

----       (1995). The Minimalist Programme. New York, Harper and row        publishers.         New York,
Evanston and London

Chomsky, N. and Halle, m. (1968). The sound patterns of English. New York, Harper and          Row
Publishers. New York, Evanston and London.

Kouega, J. (2000). From graphemes to English speech: a handbook of sounds and          spoken text
            production, vol. 1. Edition saragraph, Cameroon.

Sadembouo, E. (1980). Critere d'identification du dialecte de reference standard.University of Yaounde,
            Departement des langues Africaines et linguistique. Yaounde: unpublished Ph.D disertation.

Simons, G. F. (1994). Principles of multidialectal orthography design*: notes on literacy.   Cambridge: CUP.
Britain.*

Simons and Fennig (2018, 2020, 2022). Ethnologue: languages of the world (vol. 19). (G. F. Lewis, ed.) Dallas:
            SIL International.

Swadesh. M. (1952). Toward greater accuracy in lexicostatistic dating: *International        Journal of American
linguistics, 21, 121-137: USA.*

Seguin, L. (2004). Survey on a Shoestring. SIL and UTA. SIL International.

Terry, S. (1980). Segmental phonology of nouns in Yamba. Calgary, Canada: University of Calgary. (ed.)

Wega, S. (2004). Etude Phonologique du Polri. Master Dissertation. The University of      Yaounde 1.