
A Lexicostatistical Analysis of Nine Aari Dialects¹

Melkeneh Seid²

Yohannes Adigeh³

Abstract

This article deals with a lexicostatistical analysis of the nine Aari varieties: *Gayl, Sido, Woba, Layda, Biyo, Shengama, Baaka, Kure and Kaysa*. The right names and numbers of Aari dialects or language varieties have been describing in different ways in many past literatures and again there were no thorough studies on the mutual intelligibility level of all the language varieties of Aari. In order to show an overall intelligibility level between language varieties, lexical, phonological and grammatical (morphological and syntactical) comparisons are needed. Therefore, to fill the gap, in this article a lexical comparisons among the nine Aari language varieties are discussed. The lexical comparisons were done in three different categories (completely similar, partially similar and completely different vocabularies) based on 324 lexical items from each language variety. As a result, depending on percentages of completely similar shared lexicons the varieties are ranked from the highest to the least as *Layda, Shengama, Biyo, Kure, Woba, Baaka, Sido, Kaysa and Gayl*. In the second partially similar category the varieties are ranked from the highest shared percentage to the least as follows: *Gayl, Sido, Woba, Baaka, Biyo, Kure, Kaysa, Shengama and Layda*. In the third and final category of unshared vocabularies percentages, the language varieties are ranked from the highest unshared or dissimilar variety to the least unshared one based on their dissimilarity level

¹ This article is extracted from the monograph submitted to the Academy of Ethiopian Languages and Cultures supported by Christensen Fund.

² A PhD candidate in the department of Linguistics, Addis Ababa University

³ Assistant professor in Linguistics (*PhD*), Academy of Ethiopian Languages and Cultures, Addis Ababa University.

as follows: *Gayl, Kaysa, Sido & Baaka, Woba, Kure, Biyo, Shengama and Layda*. Finally, in the total RPV (a rank position value) result of completely similar vocabularies, based on their similarities language varieties are clustered in four groups from highest to the least one. Here, the clustering shows that lexically *Layda* language variety is considered as the center for the rest of all varieties. In contrast, the varieties *Kaysa and Gayl* have the least shared lexicons than the rest seven varieties or lexically they are too different than all the rest Aari varieties.

1. Introduction

1. 1 Background of the project

The People

All the Aari people who inhabit highland areas all are agriculturalists. They cultivate different crops and cash crops such as cardamom and coffee. Besides, they rear different domestic animals. According to the information obtained from the office of the Population and Housing Census (2008), the total population of the Aari district is about 280,187. Out of this total number of population, about 212,389 people live in South Aari district whereas the remaining 67,798 people live in North Aari district.

Melkenh (2014), described the Aari people especially those who live in south Aari Woreda are classified into three main clans. These are Indi, Amen and Ashenda. Further, these three clans are categorized under one proto-clan named Qensa [*qəntsa*]. In addition, there is another clan called Mena. Mena has three sub-divisions. The *Mena* clan does not fall under the *Qensa* clan "deemed pure Aari" because, traditionally, they are considered to be of a "lower class." Therefore, clans such as *Indi, Amen and Ashenda* are all considered themselves as the upper class clans of the Aari people.

Location

The Aari people inhabit the south-western part of Ethiopia in the Southern Nations Nationalities and Peoples Regional State. They live in the South Omo Zone, which was previously called *Geleb and Hamer Bako Awrajja*. The administrative town of this zone is called Jinka it is 781 km far from Addis Ababa, the capital city of Ethiopia and 525 km away from Hawassa, the capital city of the Southern Nations, Nationalities and Peoples Regional State. Recently, the Aari people are specifically found into two different districts (*woredas*): North (semen) Aari and South (dehub) Aari. The administrative towns of the districts are Gelila and Gazer respectively. The South Omo Zone comprises about eight districts all together. Again, the other thing which makes this zone unique from the rest of the zones in the Region is the fact that it is made up of sixteen ethnic groups out of the fifty-six ethnic groups in the SNNPRS.

The Language

As stated by Bender (1989) and Ford (1985), Aari consists of ten identified dialects: Bako, Biyo, Gelila (Gayl), Layda, Seyki, Shengama, Sido, Wubamer, Zedo and Kure. However, “The main dialect division is between Galila in the north and Wubamer in the east, separated by Bargeda valley” (Bender 1989:2). Ford (1985) argues that there is a high degree of mutual understanding among the people of these dialect areas. But Sido is the common denominator for all dialects of Aari. Furthermore, the word Aroid “refers to a language group of south western Ethiopia, which includes Aari, the Hamar cluster (Hamar, with Bena and Karo) and Dime and forms a sub-branch of the Omotic language family. Bender first coined this name in the 1970s. It has also been called South Omotic or Eastern Omotic” (Tsuge 2003: 350).

1.2. Statement of the Problem

In previous studies most researchers put the dialects of Aari in different ways. The variation is in number and name of the identified dialects.

Therefore, the classification of the Aari dialects has been a problematic issue. As stated by Bender (1989) and Ford (1985), Aari consists of ten identified dialects: *Bako*, *Biyo*, *Gelila (Gayl)*, *Layda*, *Seyki*, *Shengama*, *Sido*, *Wubamer*, *Zedo* and *Kure*. However, “The main dialect division is between *Galila* in the north and *Wubamer* in the east, separated by Bargeda valley” (Bender 1989:2). Yet again, Ford (1985) argues that there is a high degree of mutual understanding among the people of these dialect areas. But *Sido* is the common denominator for all dialects of *Aari*. Furthermore, Fekede (2011:5) has dialect lists *Gayl*, *Bargid*, *Ub*, *Sido*, *Shangama*, *Layda*, *Biyo*, *Bakka* and *Wubamer*. Fekede came across with some unfamiliar dialects such as *Ub* and *Bargid*.

Unlike the other dialects of Aari, *Gayl* is a unique and easily identifiable dialect by other dialect speakers of the language. As stated by Fekede (2011), it seems that *Gayl* deviates in its lexicon from the rest dialects of Aari. The maximum shared vocabulary *Gayl* has is 62.5 % with *Wuba* followed by 50.5% with *Sido* and list shared vocabularies 44.7% with *Bakka*. On the other hand, based on Ford, if we exclude *Galila* and *Seyki* all the other dialects have at least 85% shared vocabulary. Even so, *Galila* still has 67.5 – 79.5% shared vocabulary with others (Tsuge, 2006).

1.3. Significance of the Study

The significance of this study is much different from previous studies of the language because as stated above most previous studies highly focused on linguistic descriptions of Aari. This research, however, aims to fill the gap by providing a clear and complete lexicostatistical analysis of Aari varieties. And, it will be a big reference for other researchers and the society too.

1.4. Objectives

This article is designed to study the dialects of Aari, one of the least studied Omotic languages.

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

- Giving a detailed lexicostatistical analysis of its varieties
- Determining the appropriate names and numbers of each variety based on the lexicostatistical analysis.
- Showing intelligibility level and clustering its varieties based on their percentage and RPV result of lexical similarity and difference.

1. 5. Methods of the Study

1.5.1. Selection of Samples

The Aari language has different dialects. The people are living in several *Kebeles* in two districts called *North (semen) Aari* and *South (Debub) Aari*. For this study, the researchers have selected all the dialect areas and targeted on some *Kebeles* in north and south Aari districts. The lexical data was collected from nine sites or *Kebeles*: *Gelila, Geza, Baako, Mes'er, Shekamer, Kure, Wubhamer, Woset and Kaysa* which respectively represent the following respective Aari varieties: *Gayl, Layda, Baaka, Sido, Biyo, Kure, Woba, Shengama and Kaysa*. All the varieties are spoken in *Southern Aari District* excluding the *Gayl* variety which is found in and around *Gelila, North Aari District*.

1.5.2. Techniques Used for Collecting and Analyzing the Data

Three up to four informants from each variety were selected in the study and standard fieldwork practices were employed in collecting the data. A questionnaire was used to elicit linguistic data and linguistic elicitation session was conducted with the speakers, individually and in groups. The lexical data includes words or certain semantic fields, nomenclature (e.g. plant and animal names, kinship terms,) numerals, idioms and proverbs and cultural artifacts. The collected data were transcribed by using IPA symbols, together with English translation for each line. The linguistic data were audio recorded and transcribed phonemically and analyzed and grouped into lexical lists. The 324 lexical lists were compared using descriptive statistics, mainly percentages. Finally, a rank position value was

planned to cluster the varieties from the highest amount of shared items to the least amount and again, to cluster the varieties from the highest amount of unshared items to the least amount.

2. Conceptual Frameworks

2.1. Dialect versus Language

As language is a social phenomenon it is natural to assume that the structure of a society has some impact on the language of the speakers of that society. The impact of the structure of the society yields varieties of language use within the same language. Sociolinguistics is therefore responsible to study the causes, factors, relationships and distributions of these speech varieties. Any variety of a language characterized by systematic differences in pronunciation, grammar, and vocabulary from other varieties of the same language is called a **dialect** (Chambers and Trudgill 1998:3). **Dialectology** is the study of **regional dialects**, or dialects defined by geographical regions. Let alone at regional level, even an individual may differ in his/her speech from the community he/she belongs to what linguists call it *idiolect*.

Much of the confusion over the language/dialect distinction today derives from the ambiguities inherent in that original situation (Haugen 1966, Hudson 1996). For instance, mutual intelligibility was not considered, and neither was there a consistent divide between the two concepts of *dialect* and *language* (Harris 1990).

In previous times, “a dialect was considered as a substandard, low-status, often rustic form of language, generally associated with the peasantry, the working class and other groups lacking prestige. It was also assumed that a term is often applied to forms of language, particularly those spoken in more isolated parts of the world, which have no written form” (Chambers and Trudgill 1998: 3).

However, linguists are unanimously agree on one point: no one language is intrinsically better than any other language (Groves 2008). Focussing only

on structural features leads the linguist to have a technical, neutral definition of *dialect*. Most commonly in academic literature, *dialects* are therefore simply different but related forms of the same *language*. They are usually mutually intelligible regional or social varieties, differing in lexical, phonological, syntactic, and/or semantic ways (Wolfram 1997, Burton 2007).

2.2. Mutual Intelligibility

The common criterion to differentiate a language from dialect is a mutual intelligibility test. Accordingly, “when varieties of languages become mutually unintelligible, then they are classed as different languages. If they are mutually intelligible, they can be classed as dialects of the same language” (Groves 2008:8). Though mutual intelligibility test works for the majority of cases, there are linguistic and non-linguistic factors that affect the results of intelligibility test.

2.2.1. Political and cultural history

Mutual intelligibility can be affected by the political and cultural histories of the speakers of the language varieties in a linguistic community. For instance, Danes, Norwegians and Swedes can all understand one another, yet their varieties are considered separate languages because they are spoken in different countries. Similarly, Hindi and Urdu, and Serbo-Croatian and Bosnian are listed as separate languages for political or religious reasons, yet they are mutually intelligible (in their spoken forms) (Crystal 2000, Wardhaugh 2000). On the contrary, the Chinese language which has many hundreds of millions of speakers who are mutually unintelligible is considered as a single language.

There are many languages with different varieties which are mutually unintelligible but counted as one language because of the political ideology they follow. Therefore, the status of a language variety either to be counted as dialect or language is determined by the political or cultural history of the speakers of the varieties not by its linguistic nature. For instance, in

Ethiopia, Wolayta, Gamo, Gofa and Dawro do speak dialects of the same language but listed as different languages. The speakers of these language varieties are mutually intelligible. Similarly, Kafi Noonoo and Shaki Noonoo are registered as different languages by the government of Ethiopia though they are mutually intelligible. The issue is the same between Sil'te and Eastern Gurage.

2.2.2. The sliding scale of mutual intelligibility

It is linguistically very difficult to making a clear boundary between the dialects and concluding that such and such languages are dialects and these varieties are separate languages. Sometimes the status of mutual intelligibility is determined by the settlement of the speakers of the language. It implies that the nearby speakers of dialects of a language do understand each other but do not where they are far apart. For instance, “the speakers in one place can understand the dialect of those nearby. However, the languages of the speakers at extreme ends of the continuum are different so that they have become mutually unintelligible to each other” (Groves 2008:10). The speakers of these language varieties in between are more or less intelligible each other as per their geographical intimacy. The degree of understanding between the dialects depends on the nearness or farness of the dialects. “Based on linguistic factors alone, it is impossible to decide where these boundaries should be; political boundaries have to suffice” (Trudgill 1974, Petyt 1980, Francis 1983, Hudson 1996, Chambers and Trudgill 1998).

The difference of the language varieties is varies from place to place in the continuum. It begins from slight sound system or pronunciation difference to basic morphological or syntactic variations. In this case Chinese language can be good example. As quoted in Wang (1997:56) Lü's, comments on: Everyone knows that Chinese has many dialects, but how many are there? If slight differences in pronunciation are the basis for distinguishing dialects, then the dialects are indeed numerous.... If we require differences in the sound system, then perhaps there are many hundreds of dialects, perhaps

one or two thousand. But if the requirement is agreement on several key features, not considering other differences, then possibly there are some eight or ten dialects.... The fact we come up with seven groups is in large part an artefact of our expectations, based on linguistic as well as extra-linguistic factors.

The language varieties in Gurage cluster do have similar features; i.e. the neighbouring language varieties do manifest similar linguistic behaviours where as varieties in a distant geography are almost mutually intelligible. For instance, the degree of intelligibility between Chaha and Eza is different from Chaha and Endegagn. Similarly, the degree of intelligibility between Endegagn and Ennemor significantly varies from the relationship between Endegagn and Sodo.

2.2.3. The difficulty of mutual intelligibility measurement

How to measure mutual intelligibility significantly matters in judging the dialect versus language relationship. Sociolinguists usually try to see the status of language varieties whether they are dialect or language by asking how much the speakers of the varieties understand each other. The answer will be subjective for its difficulty of quantitative result of mutual intelligibility. It is very difficult to measure who understand what. We do not have objective instrument to measure the level of understanding among the speakers of different varieties in a certain language. At earlier times, linguists employ serious measurements of mutual intelligibility that focuses on participant intelligibility. “Two speakers talked together (or one speaker listened to another), and the percentage of understood content was taken as the degree of mutual intelligibility” (Cheng 1992). Then, if more than 50% of the content was understood, their speech varieties were considered dialects rather than separate languages (Mair 1991).

However, according to Groves (2008:11), this method does not work for all cases. Firstly, it would be very difficult to get the cut-off point to quantify in percents. If we ask someone how much percent does he/she understand the variety of a certain language, his or her answer will be based on simple

guess. Secondly, unidirectional intelligibility has to be allowed for. Mutual intelligibility is not always reciprocal, sometimes for language reasons, but more often for ‘people reasons.’ Assume that two interlocutors of different varieties talked together and be rated for their intelligibility. The speaker of variety A may better understand variety B but the speaker of variety B may not understand variety A. The degree of their intelligibility varies from direction to direction. Thirdly, the differences of personal experiences bring different results on the degree of intelligibility. The one who has better exposure for wider community and extrovert behavior may understand better than the one who has less exposure and introvert behavior. Fourthly, the dialectal variation is manifested at phonological, morphological, lexical and grammatical levels. Which one is the best base for calculation and analysis? Groves (2008:11) raises the following questions: “For instance, for vocabulary, should it be by whole-word correspondence, or phonological similarity within words? And if the latter, then what weighting should be applied to each feature? And which words should be selected for analysis?”

2.2.4. Willingness to understand

Mutual intelligibility is not between language varieties but between the speakers of these varieties (Hudson 1996:35–36). The analysis will be laid on the answer of individuals or peoples of those varieties. If we lay our ground on people, we should consider the socio-historical relationship among the communities under investigation. Differing motivation levels

between two speech communities could lead to a situation where speech community A claims that they understand the language of speech community B, while speech community B states that they do not understand A—or, possibly, refuses to try to understand them due to social and/or political reasons (Hudson 1996, Chambers and Trudgill 1998). This means the intelligibility of the two language varieties is based on the attitude of the speakers towards the speech community.

2.2.5. The educational level of the speakers

The degree of mutual intelligibility between speakers of different dialect groups can also depend upon the educational level of the speakers and upon the subject being discussed (Cheng 1992). Those who do not have good knowledge and experience on the topic being discussed tend to have a limited basic vocabulary, however, those who are educated and exposed to various linguistic communities may understand other varieties better than the illiterate one. The one who is not mobile in his/her life will have little chance for exposure to other dialects beyond their own group. By contrast, most educated persons would be able to speak some additional vocabularies and linguistic expressions other than their own variety. Therefore, the degree of intelligibility between the speakers of the dialects depends on the level of academic background of the speakers.

2.2.6. The influence of the standard written language

Two individuals from two different dialects (one is a standard written language and the other is not) do not equally understand each other. The one from a non-standard one do have a chance to know the norm of the standardized one via various media so that he/she can have an exposure for the features of the language but in contrary, the one who belongs to the standardized one may not know the linguistic behavior of the non standard dialect. For instance, the Addis Ababa dialect of Amharic is intelligible for all other Amharic dialect speakers but the speaker of Addis Ababa variety does not understand the dialects of Amharic such as Gojjam or Wollo.

In a diglossic situation a language variety introduced into the society via written form usually receives higher prestige than a variety only in oral form. The differences of the status of languages lead to a clear demarcation between the functions of registers or styles. Therefore, the one which receive better status is used for formal communication domains where as the variety which is given low status is used only for informal domains such as at home or market (Lai 2001).

In this paper the dialects of Aari are discussed based on their lexical similarity and difference. The data yielded about nine dialects.

3. Data Analysis: Results of Lexical Comparison

For the purpose of lexical comparison 324 lexical items were collected from word categories such as: nouns, adjectives, pronouns, adverbs and verbs. This lexical data collected from each language varieties were phonemically transcribed and a single language variety or dialect was compared with the rest eight varieties under Aari language. Then, the lexical data were categorized under three groups based on their similarities and differences. These three groups are: completely similar (completely shared) (CS), partially similar (shared) (PS) and completely different (CD).

If the words have similar consonant and vowel phonemes, they are considered as completely similar or completely shared. For example, *nuki* 'nose' is known as the same form in Gayl, Sido, Woba, Layda, Biyo, Shengama, Baaka, Kure, and Kaysa. Therefore, the word *nuki* is completely shared with in all nine varieties.

If half parts of the consonants or vowels of a word are shared among language varieties, we call it partially shared or similar. For instance, [*gufi*] 'navel' in Sido and Woba is known as [*gura*] in Gayl and [*guubi*] in Layda, Biyo, Shengama, Baaka, Kure and Kaysa. The words [*gufi*, *gura* and *guubi*] have the same meaning but vary in one or two speech sounds. So, they can be called partially shared or partially similar vocabularies. Similarly, [*tf'äfta*] 'rotten' in Woba and Baaka is pronounced as [*tf'äfta*] in

Biyo, Shengama and Kaysa, [*tf'äftanda*] in Kure, [*tf'äftanda*] in Layda and [*tf'affanda*] in Gayl. These five vocabularies vary in one or two or three sound systems but similar in other phonemes so it is said to be partially similar or shared.

On the other hand, if words have different consonant and vowel systems for the same meaning, they are labeled as completely different; for instance the

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

meaning of ‘gourd’ is [*pufta*] in Sido and Woba, [*oha*] in Gayl, but [*gusi*] in Baaka, Layda, Biyo, Shengama, Kure and Kaysa. These three words represent one and the same meaning so we can call it completely different vocabularies. Below, the dialects of Aari are analyzed to what extent they are lexically shared each other in percentage.

3.1. Completely Shared Vocabulary

Here we see to what extent the Aari dialects completely share their vocabularies. Most dialects differently share their words with the other dialect. In other words, each dialect shares different number of lexical items with the other dialects. Out of 324 lexical items, the Gayl dialect shares 33.6% similar lexical items with Sido, 34.6% with Woba, 26.8% with Layda etc. The percentage of completely shared vocabularies between dialects are shown in (1) below.

<u>Sido</u>	<u>Woba</u>	<u>Layda</u>	<u>Biyo</u>	<u>Sheng.</u>	<u>Baaka</u>	<u>Kure</u>	<u>Kaysa</u>	
33.6	34.6	26.8	28.4	28.1	25.9	26.5	21.6	<u>Gayl</u>
	71.9	52.8	50	49.4	50.3	52.2	40.4	<u>Sido</u>
		58.6	57.4	57.4	50.3	55.9	42.9	<u>Woba</u>
			70.9	68.2	60.2	67	54.6	<u>Layda</u>
				64.2	58.9	66.7	53.4	<u>Biyo</u>
					60.5	65.7	57.4	<u>Sheng.</u>
						61.7	53.7	<u>Baaka</u>
							52.5	<u>Kure</u>

1. Percentages of Completely Shared Vocabularies between eight dialects of Aari

The first row of chart 1 shows the extent to which the Gayl dialect has completely similar vocabularies with the other eight dialects. And the final column shows the extent to which Kaysa shares completely similar vocabularies with the other eight dialects so that it shares 21.6% with Gayl, 40.4 % with Sido etc.

To investigate the share level of each dialect, we can refer to the intersection of the dialects and see the percentile of their similarity. In this case, Sido has the highest completely shared vocabularies with Woba (71.9%) and the least completely shared vocabularies with Gayl (33.6%) and 40.4% with Kaysa. Again, Sido has nearly half or more completely shared vocabularies with the rest of the languages such as Shengama (49.4%), Biyo (50%), Baaka (50.3%), Kure (52.2%) and Layda (52.8%).

Woba shares 71.9% with Sido, 58.6% with Layda, 55.9% with Kure, 57.4% with Shengama, 50.3% with Baaka, 42.9% with Kaysa and 34.6 % with Gayl. Accordingly, the nearest Aari variety for Woba is Sido and the furthest is Gayl.

Layda shares from the highest to the least, Biyo (70.9%), Shengama (68.2%), Kure (67%), Baaka (60.2%), Woba (58.6%), Kaysa (54.6%), Sido (52.8%) and Gayl (26.8%). All language varieties have more than fifty percent completely similar vocabularies of Layda except Gayl, which shares only 26.89%.

Biyo shares 28.4% completely similar vocabularies with Gayl, 50% with Sido, 53.4% with Kaysa, 57.4% with Woba, 58.9% with Baaka, 64.2% with Shengama, 66.7% with Kure and 70.9% with Layda. Layda has the highest share of similar vocabularies with Biyo but Gayl has the least share with Biyo. The rest language varieties have an average and more than an average share of completely similar vocabularies out of the total 324 lexical items with Biyo.

Shengama is one of the highland varieties and has the following amount of completely shared vocabularies with other varieties from the highest to the lowest. Thus, Layda shares 68.2%, Kure 65.7%, Biyo 64.2%, Baaka 60.5%, Woba & Kaysa 57.4, Sido 49.4% and Gayl 28.1%. Shengama words are highly shared with most language varieties except with Gayl its similarity is only 28.1%. Woba and Kaysa have the same number of shared vocabularies with Shengama.

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

Baaka shares completely similar vocabularies, from the lowest to the highest, 25.9% with Gayl, 50.3% with Sido & Woba, 53.7% with Kaysa, 58.9% with Biyo, 60.2% with Layda, 60.5% with Shengama and 61.7% with Kure. Sido and Woba language varieties share equal percentage of completely similar vocabularies with Baaka.

Kure shares with 67% with Layda, 66.7% with Biyo, 65.7% Shengama and 61.7% Baaka. These varieties shared almost with equivalent number of completely similar vocabularies. The other varieties do also share completely similar vocabularies with Kure: such as 55.9% with Woba, 52.5% with Kaysa and 52.2% with Sido. Lastly, Kure has the least share of completely similar vocabularies with Gayl which is 26.5%.

Finally, Kaysa shares 21.6% completely similar vocabularies with Gayl, 40.4% with Sido, 42.9% with Woba, 52.5% with Kure, 53.4% with Biyo, 53.7% with Baaka, 54.6% with Layda and 57.4% with Shengama.

Generally, out of the total 324 lexical items, the highest shared percentage of completely similar vocabularies from all language varieties is between Sido & Woba, which is 71.9%. In contrary, the smallest number of completely shared varieties is between Gayl and Kaysa, which is 21.6%.

3.2. Partially Similar Vocabularies

The percentages of partially shared vocabularies of the nine language varieties are shown in (2) next page:

<u>Sido</u>	<u>Woba</u>	<u>Layda</u>	<u>Biyo</u>	<u>Sheng.</u>	<u>Baaka</u>	<u>Kure</u>	<u>Kaysa</u>	
36.7	36.7	45.1	42.9	43.2	42	44.4	42.3	<u>Gayl</u>
	18.5	35.5	40.4	38.6	33.3	36.1	36.1	<u>Sido</u>
		30.5	31.5	30.9	31.5	32.1	34.2	<u>Woba</u>
			23.1	26.2	27.8	26.2	25.6	<u>Layda</u>
				28.4	28.4	26.2	25.6	<u>Biyo</u>
					28.4	25	23.6	<u>Sheng.</u>
						26.8	30.9	<u>Baaka</u>
							26.8	<u>Kure</u>

2. Percentages of Partially Similar Vocabularies

Gayl has nearly the same shared percentage of partially similar vocabularies with six language varieties for instance 45.1% with Layda, 44.4% with Kure, 43.2% with Shengama, 42.9% with Biyo, 42.3% with Kaysa and 42% with Baaka. Similarly, Gayl has absolutely equal and the least share of partially similar vocabularies with Sido and Woba, that is 36.7%.

The Sido language variety has the least share of partially similar vocabularies with Woba (18.5%) but better to remember here that they have highest share of completely similar words than the remaining language varieties. In contrary, it has the utmost share of partially similar vocabularies with Biyo (40.4%). Besides, Sido has relatively medium and equal share of partially similar vocabularies with most language varieties such as with Shengama (38.6%), with Woba (36.7%), with Kaysa and Kure (36.1%), with Layda (35.5%) and with Baaka (33.3%).

Except Sido, which has the least shared of partially similar words with Woba (18.5%), Woba has nearly the same share of partially similar vocabularies with the rest Aari varieties for instance 36.7% with Gayl, 34.2% with Kaysa, 32.1% with Kure, 31.5% with Baaka and Biyo, 30.9% with Shengama and 30.5% with Layda.

Layda shares partially similar vocabularies with all the eight varieties such as, from the highest to the least, 45.1% with Gayl, 35.5% with Sido, 30.5% with Woba, 27.8% with Baaka, 26.2% with Kure & Shengama, 25.6% with Kaysa and 23.1% with Biyo.

The maximum share of partially similar vocabularies of Biyo is 42.9% with Gayl and it has the next high partially similar share with Sido (40.4%) and 31.5% with Woba. Again, nearly the same percentage (28.4%) of partially similar vocabularies has with Baaka and Shengama, 26.2% with Kure, 25.6% with Kaysa and 23.1% with Layda.

Shengama shares the least partially similar words with Kaysa (23.6%) and 25% with Kure, which ranks the second from the last. Shengama's the third

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

least shared percentage of partially similar vocabularies is 26.2% with Layda. Again, it has completely equal share of partially similar words with Baaka and Biyo (28.4%). The top three shared percentage of partially similar words of Shengama are 43.2%, 38.6% & 30.9% with Gayl, Sido and Woba respectively.

Baaka shares 42% partially similar vocabularies with, Gayl, 33.3% with Sido, 31.5% with Woba, 30.9% with Kaysa, 28.4% with Shengama and Biyo, 27.8% with Layda and 26.8% with Kure, from the highest to the least.

Kure shares highly 44.4% partially similar vocabularies with Gayl, 36.1% with Sido, 32.1% with Woba and relatively has a medium and similar share with Kaysa and Baaka (26.8%), with Layda and Biyo (26.2%). The least shared percentage of partially similar words of Kure is 25% with Shengama.

Kaysa shares partially similar vocabularies, from the least to the highest, 23.6% with Shengama, 25.6% with Layda and Biyo, 26.8% with Kure, 30.9% with Baaka, 34.2% with Woba 36.1% with Sido and 42.3% with Gayl. Out of all varieties, the highest shared percentage of partially similar vocabularies is 45.1% which is between Gayl and Layda and the least shared percentage is 18.5% between Sido and Woba.

3.3 Percentages of Completely Different Vocabularies

The percentages of completely different vocabularies of the Aari language varieties are presented in (3) below.

<u>Sido</u>	<u>Woba</u>	<u>Layda</u>	<u>Biyo</u>	<u>Sheng.</u>	<u>Baaka</u>	<u>Kure</u>	<u>Kaysa</u>	
29.6	28.7	28.1	28.7	28.7	32.1	29	36.1	<u>Gayl</u>
	9.6	11.7	9.6	12	16.4	11.7	23.4	<u>Sido</u>
		10.8	11.1	11.7	18.2	12	22.8	<u>Woba</u>
			5.9	5.5	12	6.8	19.7	<u>Layda</u>
				7.4	12.6	7.1	21	<u>Biyo</u>
					11.1	9.2	18.8	<u>Sheng.</u>
						11.4	15.4	<u>Baaka</u>
							20.8	<u>Kure</u>

3. Percentages of Completely Different Vocabularies

The highest percentage of completely different vocabularies of Gayl is 36.1% with Kaysa and 32.1% with Baaka. Sido and Kure have 29.6% and 29% completely different vocabularies, respectively with Gayl. The Gayl variety has 28.7% completely different vocabularies with the three Aari varieties: Shengama, Biyo and Woba, and Gayl's the least unshared vocabularies percentage is 28.1% with Layda.

The percentages of completely different vocabularies between Aari varieties are as follow: Sido has 9.6% completely different vocabularies from Biyo and Woba, 11.7% from Kure and Layda, 12% from Shengama, 16.4% from Baaka, 23.4% from Kaysa and 29.6% from Gayl.

The biggest lexical difference of Woba is 28.7% with Gayl variety. Woba has 22.8% completely different vocabularies with Kaysa, 18.2% with Baaka. The minimum completely different vocabularies percentage of Woba is 9.6% which is with Sido. Other Aari varieties such as Layda, Biyo,

Shengama and Kure have nearly the same degree of difference with Woba, i.e., 10.8%, 11.1%, 11.7% and 12%, respectively.

The degree of difference between Layda and other varieties is as follows: The highest difference is seen between Gayl which is 28.1%. And it differs 19.7% from Kaysa, 12% from Baaka, 11.7% from Sido, and 10.8% from

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

Woba. On the other hand, Layda has nearly the same number of unshared vocabularies with Shengama (5.5%), Biyo (5.9%) and Kure (6.8%).

Biyo has 5.9% completely different vocabularies with Layda, 7.1% with Kure, 7.4% with Shengama, 9.6% with Sido, 11.1% with Woba, 12.6% with Baaka, 21% with Kaysa and 28.7% with Gayl, from the smallest to the biggest figure.

Out of the 324 basic words, Shengama differs by 5.5% vocabularies from Layda, 7.4% from Biyo and 9.2% from Kure. Shengama has 28.7% and 18.8% completely different vocabularies from Gayl and Kaysa, respectively. Shengama has almost equivalent degree of differences from Aari varieties such as 11.1% from Baaka, 11.7% from Woba and 12% from Sido.

The calculated degree differences of unshared words between Baaka and the other Aari varieties are shown from the biggest to the smallest percentage as follow: 32.1% with Gayl, 18.2% with Woba, 16.4% with Sido, 15.4% with Kaysa, 12.6% with Biyo, 12% with Layda, 11.4% with Kure and 11.1% with Shengama.

Kure has the least percentage of unshared vocabularies with Layda (6.8%), Biyo (7.1%) and Shengama (9.2%). The other Aari varieties like Baaka, Sido and Woba has 11.4%, 11.7% and 12% completely different vocabularies with Kure respectively. On the other hand, Gayl and Kaysa have the highest completely different vocabularies with Kure, which is 29% & 20.8% respectively.

The highest lexical difference of Kaysa variety is 36.1% which is observed between Gayl. Kaysa differs by 23.4% from Sido, 22.8% from Woba, 21% from Biyo, 20.8% from Kure, 19.7% from Layda and 18.8% from Shengama. The least difference is 15.4% between Kaysa and Baaka. In general, the highest percentage of unshared vocabularies from all Aari varieties is 36.1%, which is between Gayl and Kaysa. In contrary, the least is 5.5% which is between Sido and Woba varieties.

3.4 Clustering the Varieties of Aari

In an attempt to find the cross-dialect relationships between the Aari varieties, a rank position value (RPV) is worked out. Primarily, the Aari varieties are ranked based on their percentage values as 1st, 2nd, 3rd ... 9th. Then after, RPVs are allocated. The compared Aari language varieties are nine in number. Then, we shall give the highest RPV 9 for the target language variety presuming a 100% likeness when it is compared with itself. The language variety, which has the highest percentage of similarity with the target language is given the RPV 8 and based on percentages of similarity they have with target language, the rest language varieties are given the RPV from 7 up to 1. If two language varieties have similar rank order to one of the dialects of Aari, their rank will be added and the result is divided into two; then the two language varieties receive the result of the division. The language variety that has a rank next to the language varieties which already got the same rank passes over an immediate rank next to the rank received by both language varieties and receives its rank following the bypassed one because the language varieties which had the same rank are expected as they received successive ranks. The suggested rank orders and the rank position values (RPVs) are used to cluster the Aari varieties depend on the lexical relationship between the language varieties under investigation. Completely and partially similar shared vocabularies are taken for lexical comparison. Based on the percentages of completely similar shared vocabularies, the relational rank order of Aari varieties is shown in (4) next page.

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

Target Language	Relational Rank, high to low, from left to right							
	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th	8 th
GA	WO	SI	BI	SH	LA	KU	BA	KA
SI	WO	LA	KU	BA	BI	SH	KA	GA
WO	SI	LA	SH & BI	KU	BA	KA	GA	
LA	BI	SH	KU	BA	WO	KA	SI	GA
BI	LA	KU	SH	BA	WO	KA	SI	GA
SH	LA	KU	BI	BA	KA & WO	SI	GA	
BA	KU	SH	LA	BI	KA	WO & SI	GA	
KU	LA	BI	SH	BA	WO	KA	SI	GA
KA	SH	LA	BA	BI	KU	WO	SI	GA

4. Relational Rank based on Completely Similar Shared Vocabularies

The Relational ranks based on completely similar shared vocabularies in (4) above are converted in to RPV as in (5) below.

	GA	SI	WO	LA	BI	SH	BA	KU	KA
GA	9	7	8	4	6	5	2	3	1
SI	1	9	8	7	4	3	5	6	2
WO	1	8	9	7	5.5	5.5	3	4	2
LA	1	2	4	9	8	7	5	6	3
BI	1	2	4	8	9	6	5	7	3
SH	1	2	3.5	8	6	9	5	7	3.5
BA	1	2.5	2.5	6	5	7	9	8	4
KU	1	2	4	8	7	6	5	9	3
KA	1	2	3	7	5	8	6	4	9
Total	17	36.5	46	64	55.5	56.5	45	54	30.5

5. RPV based on completely similar shared vocabularies

Based on the completely similar shared vocabularies, we can cluster the nine Aari language varieties, from the highest to the least as: 1st(Layda), 2nd(Shengama), 3rd(Biyo), 4th(Kure), 5th(Woba), 6th(Baaka), 7th(Sido), 8th(Kaysa) and 9th(Gayl).

Depending on the percentages of partially similar shared vocabularies, the relational rank order of language varieties is shown in (6) below.

Target Language	Relational Rank, high to low, from left to right							
	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th	8 th
GA	LA	KU	SH	BI	KA	BA	WO&SI	
SI	BI	SH	GA	KU&KA	LA	BA	WO	
WO	GA	KA	KU	BA & BI	SH	LA	SI	
LA	GA	SI	WO	BA	KU & SH	KA	BI	
BI	GA	SI	WO	BA & SH	KU	KA	LA	
SH	GA	SI	WO	BA & BI	LA	KU	KA	
BA	GA	SI	WO	KA	SH & BI	LA	KU	
KU	GA	SI	WO	KA&BA	BI & LA	SH		
KA	GA	SI	WO	BA	KU	BI & LA	SH	

6. Relational Rank based on Partially Similar Shared Vocabularies

The Relational ranks based on partially similar shared vocabularies in (6) above are converted in to RPV as in (7) below.

	GA	SI	WO	LA	BI	SH	BA	KU	KA
GA	9	1.5	1.5	8	5	6	3	7	4
SI	6	9	1	3	8	7	2	4.5	4.5
WO	8	1	9	2	4.5	3	4.5	6	7
LA	8	7	6	9	1	3.5	5	3.5	2
BI	8	7	6	1	9	4.5	4.5	3	2
SH	8	7	6	3	4.5	9	4.5	2	1
BA	8	7	6	2	3.5	3.5	9	1	5
KU	8	7	6	2.5	2.5	1	4.5	9	4.5
KA	8	7	6	2.5	2.5	1	5	4	9
Total	71	53.5	47.5	33	40.5	38.5	42	40	39

7. RPV based on partially similar shared vocabularies

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

Based on the partially similar shared vocabularies, we can cluster the nine Aari language varieties, from the highest to the least as follows: 1st(Gayl), 2nd(Sido), 3rd(Woba), 4th(Baaka), 5th(Biyo), 6th(Kure), 7th(Kaysa), 8th(Shengama) and 9th(Layda).

In contrary with the relational rank based on similarity, here the rank of Aari varieties showed based on percentage of unshared vocabularies. The unrelation rank order of Aari varieties is shown in (8) below.

Target Language	Unrelation Rank, high to low, from left to right							
	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th	8 th
GA	KA	BA	SI	KU	WO, BI & SH			LA
SI	GA	KA	BA	SH	LA & KU		WO & BI	
WO	GA	KA	BA	KU	SH	BI	LA	SI
LA	GA	KA	BA	SI	WO	KU	BI	SH
BI	GA	KA	BA	WO	SI	SH	KU	LA
SH	GA	KA	SI	WO	BA	KU	BI	LA
BA	GA	WO	SI	KA	BI	LA	KU	SH
KU	GA	KA	WO	SI	BA	SH	BI	LA
KA	GA	SI	WO	BI	KU	LA	SH	BA

8. Unrelation Rank Based on Completely Different (unshared) Vocabularies

The unrelation ranks based on completely different (unshared) vocabularies in (8) above are converted in to URPV as in (9) below.

	GA	SI	WO	LA	BI	SH	BA	KU	KA
GA	x	6	3	1	3	3	7	5	8
SI	8	x	1.5	3.5	1.5	5	6	3.5	7
WO	8	1	x	2	3	4	6	5	7
LA	8	5	4	x	2	1	6	3	7
BI	8	4	5	1	x	3	6	2	7
SH	8	6	5	1	2	x	4	3	7
BA	8	6	7	3	4	1	x	2	5
KU	8	5	6	1	2	3	4	x	7
KA	8	7	6	3	5	2	1	4	x
Total	64	40	37.5	15.5	22.5	22	40	27.5	55

9. URPV based on completely different (unshared) vocabularies

Based on the completely different (unshared) vocabularies, we can cluster the nine Aari language varieties, from the highest to the least as: 1st(Gayl), 2nd(Kaysa), 3rd(Sido & Baaka), 4th(Woba), 5th(Kure), 6th(Biyo), 7th(Shengama) and 8th(Layda).

3.5. Overall lexical Relations

The overall lexical relationships between the Aari varieties based on completely similar, partially similar and completely different vocabularies are summated in (10).

(10) Relative ranks in three lexical categories

Lexical Category: **Language varieties (from high to low based on, similarity, partial similarity and difference)**

<u>Completely similar</u>	LA	SH,	BI,	KU,	WO,	BA,	SI,	KA &	GA
<u>Partially similar</u>	GA,	SI,	WO,	BA,	BI,	KU,	SH,	KA &	LA
<u>Completely different</u>	GA,	KA,	(SI & BA),	WO,	KU,	BI,	SH &	LA	

3.5.1. The RPV for completely similar and completely different

lexical categories

In order to find this, first the RPV of the partially similar vocabularies of each variety is divided in to two. Then, the result is added to the RPV's of completely shared vocabularies and completely unshared vocabularies of each language variety. Because partially similar vocabularies are on the other hand they are partially different. Therefore, partially they belong to

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

completely similar vocabularies and again partially they belong to completely different vocabularies. The RPV and their divided result of partially similar vocabularies presented in (11) below.

(11) *The RPV of partially similar vocabularies and their divided result*

RPV of partially similar vocabularies and their divided result									
	GA	SI	WO	LA	BI	SH	BA	KU	KA
Partially similar	71	53.5	47.5	33	40.5	38.5	42	40	39
RPV ÷ 2	35.5	26.75	23.75	16.5	20.25	19.25	21	20	19.5

3.5.1.1 The total RPV of Completely Similar Vocabularies

In order to get the total RPV of CSV, the above divided results of the partially similar vocabularies are added to the RPV of the completely similar vocabularies of Aari varieties as summarized in (12).

	GA	SI	WO	LA	BI	SH	BA	KU	KA	Total
CSV	17	36.5	46	64	55.5	56.5	45	54	30.5	405
+ Result of RPV of PSV	35.5	26.75	23.75	16.5	20.25	19.25	21	20	19.5	202.5
Total	52.5	63.25	69.75	80.5	75.75	75.75	66	74	50	607.5

12. *The total RPV of completely similar vocabularies*

Based on the total RPV result of completely shared or similar vocabularies, we can cluster the nine varieties, from the highest to the least as: *Layda*, (*Biyo*, *Shengama*), *Kure*, *Woba*, *Baaka*, *Sido*, *Gayl* and *Kaysa*. When we compare the above total result, some varieties are completely the same and some others are nearly the same with a minimum gap in terms of completely shared vocabularies. Therefore, we can revise the clustering or regroup them as in the following figure.

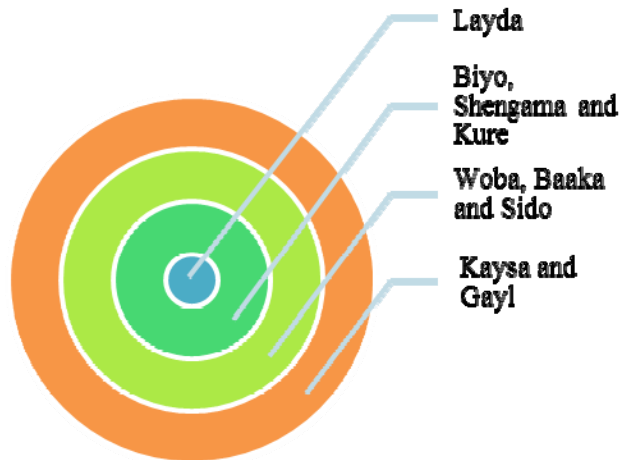


Figure 1: Cross- language variety relationship based on total RPV of completely shared vocabularies

Based on shared lexicon, the language variety in the center of the circle, *Layda* is the core of all the rest eight varieties of Aari. Language varieties in the second inner circle (*Biyo*, *Shengama* and *Kure*) are the same or nearly the same varieties based on their shared lexicon. So, they can be categorized as the second more accessible than the rest five varieties of Aari. The varieties in the third circle (*Woba*, *Baaka* and *Sido*) are the third accessible based on their shared lexicon than the varieties in the last circle (*Kaysa* and

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

Gayl). We can decide that the varieties in the outer fourth circle (*Kaysa and Gayl*) have the least shared lexicon than the rest seven varieties.

3.5.1.2. The total RPV of Completely Different Vocabularies

In order to get the total RPV of completely different (unshared) vocabularies, the divided result of partially similar vocabularies are added to the RPV of CDV of Aari varieties as summarized in (13).

RPV of completely different vocabularies and addition of the RPV's result of partially similar vocabularies										
	GA	SI	WO	LA	BI	SH	BA	KU	KA	Total
<i>CDV</i>	64	40	37.5	15.5	22.5	22	40	27.5	55	324
+ <i>Result of RPV of PSV</i>	35.5	26.75	23.75	16.5	20.25	19.25	21	20	19.5	202.5
<i>Total</i>	99.5	66.75	61.25	32	42.75	41.25	61	47.5	74.5	526.5

13. The total RPV of completely different vocabularies

Based on the above total RPV of completely different vocabularies we can cluster the Aari varieties from the highest dissimilarity to the lowest as follow: *Gayl, Kaysa, Sido, Woba, Baaka, Kure, Biyo, Shengama and Layda*. Some varieties have minimum difference on their shared percentage of dissimilar vocabularies. Hence, regrouping of cross-language varieties

based on their percentages closeness of unshared lexicon is presented as in (Figure 2) below:

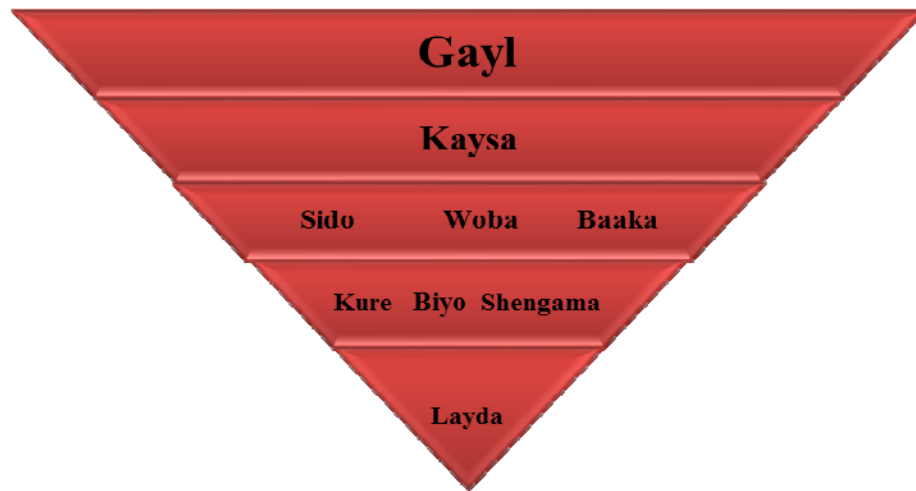


Figure 2: Cross-language varieties similarity based on total RPV of unshared (CD) vocabularies

4. Conclusion

A lexicostatistical analysis between the nine Aari varieties has been discussed in this article. The lexicons were categorized under three groups (*completely shared, partially shared and completely different*) founded on their similarities and differences. Then, RPV (a rank position value) was calculated to cluster the varieties from the highest amount of shared or unshared items to the least. The total RPV result among nine varieties based on completely shared vocabularies percentage from the highest to the least showed that *Layda, (Biyo, Shengama), Kure, Woba, Baaka, Sido, Gayl and Kaysa* respectively. Unlike the previous studies, in this study, the "unknown dialect", *Kaysa* is included. Because of the information from many Aari varieties speakers, *Kaysa* area was targeted as the one sample area to collect the lexical data. The *Kaysa* kebele is found at the gate site of *Jinka* and it is geographically bordered with the people of *Bena* language speakers. There are some more kebeles, which have the same speakers like *Kaysa* area. As a

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

result, Kaysa has the least total RPV of shared vocabularies. Hence, it is clustered with Gayl variety, which has also highest percentage of unshared vocabularies or has the second least percentage of shared vocabularies.

Various scholars have been forwarding different number of dialects for Aari since 1970s. This analysis is based on lexical comparisons between the varieties existing where Aari people dwell at different geographical settings. Hence, the researchers have attempted to elicit 324 basic lexical entries from all areas of Aari speaking communities and made analysis. Finally, the result showed that nine dialects are vividly observed. The closeness and difference among them varies from one to another. The relational rank of Aari varieties between each of them was analyzed. Accordingly, Gayl has been identified as a separate variety among others as per the calculation of the analysis. Sido takes the second rank to be a special dialect next to Gayl and followed by Woba. Contrary to Bender (1989) and Ford (1985), in this study the language variety 'Zedo' is not mentioned because the researchers couldn't find any language variety or specific place around Aari speakers living in. Fekede (2011) stated that the Gayl language variety deviates in its lexicon from the rest dialects. This study also proves that Gayl has the least shared vocabularies and as the same time it has the highest percentage of unshared vocabularies. Again, based on the total RPV of shared lexicons it is clustered together with Kaysa variety, which has the least result. Unlike Tsuge's (2006), description that Gayl has 65.5% - 79.5% shared lexicons with the rest varieties, in this study, the highest shared lexicons of Gayl is 34.6% with Woba and it has the least shared lexicon 21.6% with Kaysa. Even in partially shared lexicons, the highest percentage of Gayl is 45.1% with Layda and it has the least share 36.7% with Sido and Woba varieties. Generally, in this study about nine Aari language varieties are identified in number and namely they are: *Gayl, Woba, Sido, Biyo, Layda, Shengama, Baaka, Kure and Kaysa*. Depends on completely the same and nearly the same (having a little gap) results of the total RPV of similar vocabularies or shared lexicons between varieties, they can be regrouped or clustered in four categories from the highest to the least as follows: 1st (*Layda*), 2nd (*Biyo, Shengama and Kure*), 3rd (*Woba, Baaka and Sido*) and 4th (*Kaysa and*

Gayl). Layda shares most vocabularies with the rest eight language varieties of Aari. Therefore, lexically *Layda* language variety is considered as the center for the rest all varieties. On the other hand, the varieties in the fourth category (*Kaysa and Gayl*) have the least shared lexicons than the rest seven varieties. To sum up, this lexicostatistical comparison study is a great means or opportunity to do an overall structural (lexical, Phonological and grammatical) intelligibility among Aari varieties.

List of Abbreviations

ATP	Average total percentage
BA	Baaka
BI	Biyo
CD	Completely different
CDV	Completely different vocabularies
CDV	Completely different vocabularies
CS	Completely similar
CSSV	Completely similar shared vocabularies
CSV	Completely similar vocabularies
GA	Gayl
KA	Kaysa
KU	Kure
LA	Layda
PCDV	Percentages of completely different vocabularies
PS	Partially similar
PSSV	Partially similar shared vocabularies
PSV	Partially similar vocabularies
RPV	Rank position value
SH	Shengama
Sheng.	Shengama
SI	Sido
URPV	Unshared relational position value
WO	Woba

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

References

- Bender, ML. 1976. *Language in Ethiopia*. Oxford University Press. London.
- Bender, ML. 1989. “Comparative Aroid (South Omotic) Morphology” Paper prepared for 2nd *International Symposium on Cushitic and Omotic Nov. 16-18, Turin, Italy*.
- Burton, Philip (2007). *Working with Dialect Translations*. Retrieved on 31 June 2017 from <http://www.hca.heacademy.ac.uk/resources/classics/coll04/Burton.pdf>.
- Chambers, J. K., and Peter Trudgill (1998). *Dialectology* (2nd ed.). Cambridge: Cambridge University Press.
- Cheng, Chin-Chuan (1992). Syllable-Based Dialect Classification and Mutual Intelligibility
- Crystal, D. 2000. *Language Death*. Cambridge: Cambridge University Press.
- Fekede Menuta. 2011. *Aari Language: A Comparative Grammar*. Addis Ababa: Bole Printing Enterprise.
- Ford Carolyn.1985. *Report on Aari Linguistic Survey*. Submitted to Language Academy. Ministry of Culture. Addis Ababa, Ethiopia (unpublished)
- Francis, W. N. (1983). *Dialectology*. New York: Longman Inc.
- Groves M. Julie “Language or Dialect—or Topolect?” *Sino-Platonic Papers*, 179 (February, 2008).
- Harris, Roy (1990). The Dialect Myth, in Jerold A. Edmondson, et al. (eds.) *Development and Diversity: Language Variation across Time and Space: A Festschrift for Charles-James N. Bailey*. Arlington: The Summer Institute of Linguistics, Publication 93.
- Haugen, E. (1966). Dialect, Language, Nation. Repr. in J. B. Pride and Janet Holmes (eds.), *Sociolinguistics: Selected Readings* 1972:97–111. Harmondsworth, Eng.: Penguin Books.

- Hudson, R. A. (1996). *Sociolinguistics* (2nd Ed.). Cambridge: Cambridge University Press.
- Lai, Mee-Ling (2001). Hong Kong Students' Attitudes towards Cantonese, Putonghua and English after the Change of Sovereignty, in *Journal of Multilingual and Multicultural Development* **22(2)**:112–133. Clevedon, Avon, Eng.: Tieto.
- Mair, Victor H. (1991). What Is a Chinese 'Dialect/Topolect'? Reflections on Some Key Sino-English Linguistic terms, in *Sino-Platonic Papers* 29:1–31. Philadelphia: Department of East Asian Languages and Civilizations, University of Pennsylvania.
- Melkeneh Seid (2014). A General Grammatical Description of Aari. A Christensen Fund Project Work Submitted to the Academy of Ethiopian Languages and Cultures.
- People's Democratic Republic of Ethiopia, Office of the Population and Housing Census, 2008.
- Petyt, K. M. (1980). *The Study of Dialect: An Introduction to Dialectology*. London: Andre Deutsch Limited.
- Trudgill, Peter (1974). *Sociolinguistics: An Introduction*. Middlesex, New York: Penguin Books.
- Tsuge, Yochi. 2003. 'Aari' in Uhlig, Siegbert(ed), In *Encyclopedia Aethiopica*. Vol. 1. Otto Harrassowitz. pp. 1-2.
- _____. 2006. On the Galila Dialect of the Aari Language. In *Cushitic-Omotiic Studies 2006*. A Report Submitted to the Japan Ministry of Education, Culture, Sports, Science and Technology for a Grant-in-Aid for Scientific Research (B), 2004 – 2007. Yamoguchi, Japan.
- Wang, William S.-Y. (1997). Languages or Dialects? in *The CUHK Journal of Humanities* **1**:54 – 62. Hong Kong: Faculty of Arts, Chinese University of Hong Kong.

A Lexicostatistical Analysis of Nine ... – Melkeneh and Yohannes

Wardhaugh, Ronald (2000). *An Introduction to Sociolinguistics* (3rd ed.). Beijing: Foreign Language Teaching and Research Press/Blackwell Publishers Ltd.

Wolfram, Walt (1997). Dialect in Society, in Florian Coulmas (ed.), 107–126.