

The Gbe Language Continuum of West Africa: A Synchronic Typological Approach to Prioritizing In-depth Sociolinguistic Research on Literature Extensibility

Angela Kluge
SIL International

For large language continua, such as Gbe of West Africa, a major challenge stakeholders concerned with language planning issues are facing is the question of the extensibility of written materials. To adequately inform language development program decisions in terms of literature extensibility, ideally, in-depth sociolinguistic studies would be conducted in all of the varieties involved, an undertaking that proves rather unfeasible for large continua. This paper proposes that when applying a multifaceted approach to linguistic data interpretation, such in-depth sociolinguistic studies may be unnecessary. This paper compares the findings of synchronic analyses of lexical and grammatical features elicited among 49 Gbe varieties to the outcomes of more recent Rapid Appraisal sociolinguistic surveys, and the findings of two diachronic studies. In considering the different approaches as complementing rather than contradicting each other, the proposed multifaceted approach exemplifies how different outcomes from various studies may be used to inform focused in-depth sociolinguistic research. Such research would in turn directly influence language program development decisions as to the extensibility of already existing literature to the remaining speech communities of the continuum.

1. INTRODUCTION.¹ For large language continua, such as those found in West Africa, one of the major challenges that present themselves to governments, language planners, educators, (socio)linguists, and local speech community members concerned with language-planning issues or linguistic studies, is the question of extensibility of written materials: how many varieties of the continuum, and which ones, need to be targeted for language development efforts to provide all speech communities of the continuum with adequate access to these standardization efforts? Ideally, in-depth sociolinguistic studies would be conducted in all of the varieties under investigation to adequately inform language development program decisions as to the extensibility of already existing literature to the remaining speech communities of the continuum. In the context of large language continua, however, such studies prove rather unfeasible.

In the context of an extensibility study of the Gbe language continuum (Kwa language family) of West Africa, this paper proposes that in applying a multifaceted approach to lin-

¹I would like to thank Sandra Lee of SIL International and an anonymous reviewer for their helpful comments on this paper.

guistic data interpretation, such in-depth sociolinguistic studies may not be necessary. The study reported here draws on the findings of a synchronic analysis of lexical and grammatical features elicited among 49 Gbe language varieties, and compares the findings of this analysis to the results of more recent Rapid Appraisal sociolinguistic surveys (Bergman 1991), as well as to the findings of Capo's (1991) and Stewart's (1994) diachronic studies. In considering the different approaches as complementing rather than contradicting each other, the multifaceted approach proposed here exemplifies how different outcomes from the various studies may be used to inform focused in-depth sociolinguistic research, which would in turn directly influence language program development decisions in terms of the extensibility of already existing literature for the entire continuum.

The Gbe language continuum is situated in the southeastern part of West Africa. Thus far, five Gbe varieties have undergone language-based development on a larger scale: Aja, Ewe, Fon, Gen, and Gun.² To assess whether and to what extent the remaining communities could benefit from these existing literacy efforts, or whether additional language-based development programs in some of the remaining communities would be beneficial, the Togo-Benin branch of SIL International launched a sociolinguistic study of the language continuum in the late 1980s.

Promoting the use of already developed language materials by as wide a population as possible, SIL sociolinguistic studies typically explore the extensibility of already existing materials among related speech varieties by investigating issues related to accessibility (e.g., dialect comprehension) and acceptability (i.e., ethnolinguistic identity, language attitudes, and social factors) (Gordon 2005; Kindell 1994; Simons and Bergman 2004). To provide as wide a base as possible for language development program decisions, such extensibility studies usually investigate the above-mentioned factors through a variety of test instruments such as literature research on previous linguistic and sociolinguistic studies, word and phrase list elicitation, sociolinguistic questionnaires, informal interviews, and comprehension and bilingualism testing (International Language Assessment Conference 1990).

Such a broad-based approach was chosen for SIL Togo-Benin's sociolinguistic study on Gbe. During the first phase of the study, word and phrase lists were elicited in 49 Gbe varieties, chosen on the basis of Capo's (1986) comparative study of the Gbe language continuum. Given the overall goal of SIL's larger Gbe study of exploring the extensibility of already existing materials among the remaining Gbe varieties, the main objective of the linguistic analysis was not to establish the genetic classification of the investigated language varieties, but to obtain a rough estimate how these varieties might be treated as clusters and to establish priorities for further sociolinguistic research (Kluge 2000, 2005, 2006). Therefore, a synchronic rather than a diachronic approach was chosen for the analysis of the elicited linguistic data, with no attempt being made to establish the genetic relationships of the investigated varieties.

A detailed description of the procedures and findings of these analyses has been presented in two earlier papers: "A synchronic lexical study of Gbe language varieties: The effects of different similarity judgment criteria" (Kluge 2005) focuses on the word-list analy-

² In general, all language names are spelled using the Latin alphabet and without diacritics. For a more complete listing of alternative spellings see Kluge 2000:88.

sis, whereas the phrase-list analysis is discussed in “Qualitative and quantitative analysis of grammatical features elicited among the Gbe language varieties of West Africa” (Kluge 2006). The focus of the current paper is to summarize the findings of the synchronic lexical and grammatical analyses and to explore to what extent the findings of both analyses concur and yield the same conclusions.

After a discussion of pertinent background information on the Gbe language continuum, including previous linguistic research on Gbe, in section 2, section 3 briefly describes the procedures involved in SIL’s synchronic analysis of word- and phrase lists elicited in 49 Gbe varieties. The findings of the lexical and grammatical analysis are summarized in section 4, followed in section 5 by a discussion of these results. The findings of the synchronic lexical and grammatical analyses are discussed and compared (1) to each other, (2) to the results of more recent Rapid Appraisal sociolinguistic surveys conducted among some of the Gbe communities, and (3)—although based on a synchronic rather than diachronic approach—to the findings of Capo’s (1991) and Stewart’s (1994) diachronic analyses. The main objective of these comparisons is to obtain as comprehensive an overview of the Gbe language situation as possible, and to explore whether and to what extent the results of these rather different approaches complement or contradict each other. As will be shown, for a number of Gbe varieties the different approaches result in different outcomes as to the practical clustering of these varieties. Given these disparities, a discussion follows as to how the different approaches should be weighted and what conclusions can be drawn in terms of prioritizing focused sociolinguistic research in the context of an extensibility study of a language continuum.

2. BACKGROUND INFORMATION ON GBE.

2.1 LANGUAGE AREA. The Gbe language varieties, spoken by a total of approximately eight million speakers (estimate in Gordon 2005), are situated in the southeastern part of West Africa, occupying large areas in southwestern Nigeria, southern Benin, Togo, and southeastern Ghana. Stewart (pers. comm., reported in Williamson and Blench 2000:29), revising his 1989 classification which was based on Bennett and Sterk (1977, in Stewart 1989), proposes the following classification for the Gbe language varieties:

Niger-Congo, Atlantic-Congo, Volta-Congo, Kwa, Gbe

Based on phonological and morphophonological characteristics, Capo (1991:15) suggests the internal classification of Gbe into five major branches, Aja, Ewe, Fon, Gen, and Phla-Phera, whereas Stewart (1994) proposes an alternative internal classification into two major branches, Eastern and Western Gbe. Both classifications are discussed in more detail in section 2.3.

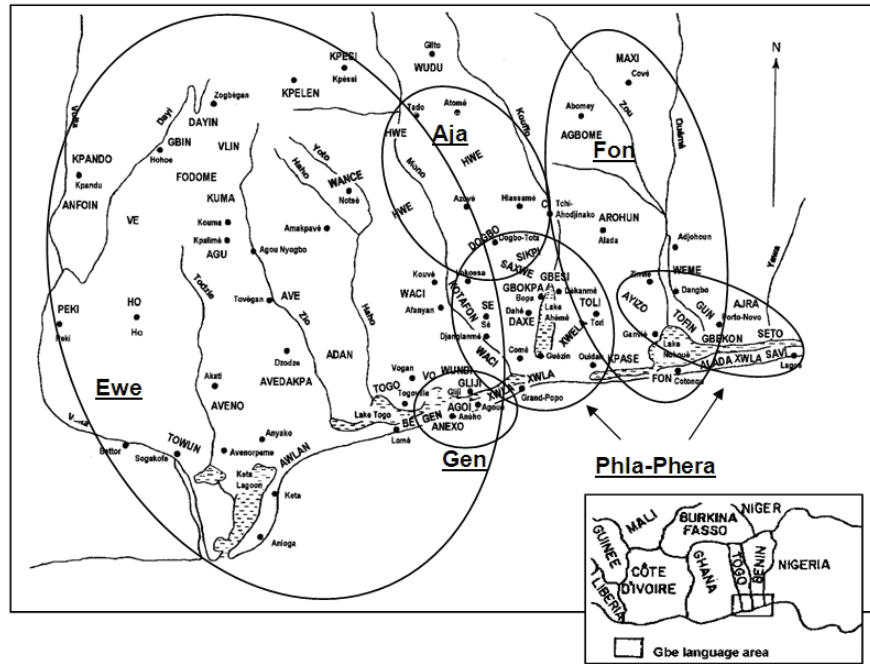


FIGURE 1: Map of the Gbe language area (based on Capo 1986: map 1a, 1991)

2.2 LANGUAGE DEVELOPMENT. Thus far, four of the five Gbe clusters suggested by Capo (1991) have been targeted for language-based development on a relatively large scale, Aja, Ewe, Fon, and Gen.

With regard to the Ewe cluster, language-based development efforts were initially based on Awlan, the most prestigious variety of the cluster (Capo 1986; Duthie 1988). These development efforts resulted in the development of standard Ewe. In Ghana, Ewe is a language of formal education for secondary school and university education, whereas in Togo, Ewe, together with Kabiye and French, is a major language of nonformal education.

In Benin, in addition to the Bariba, Dendi, Ditammari, and Yoruba languages, the Aja and Fon clusters were targeted by the Benin government in 1992 for large-scale nonformal adult education (Direction de l'Alphabétisation 1992). For the Aja cluster, no single variety has yet emerged as the standard form of Aja. However, thus far Aja-Dogbo has more published material than the remaining Aja varieties, although Aja-Hwe is the most geographically widespread variety and has an increasing body of written material in general circulation (Sullivan 2000; Gordon 2005). As regards the Fon cluster, the Fon variety³ was chosen for large-scale development efforts. In addition to the Fon variety, a second variety of the Fon cluster, Gun, was chosen for smaller-scale nonformal education in the Ouémé

³ Capo (1986) lists neither a distinct Fon nor a distinct Gen variety (see section 2.3). For the current study, however, word lists were elicited in the Fon and Gen varieties (see section 4).

region, in southeastern Benin (Direction de l'Alphabétisation n.d., a, b, c).

In addition to the Aja and Fon clusters, the Gen cluster was targeted for smaller-scale nonformal education in Benin's Mono region, in southwestern Benin, with the nonformal education efforts making use of a Waci-Gen primer which employs the vocabulary that is shared by Gen³ and Waci, another Gbe variety of Benin's Mono region (Capo 1986:44f, 102; Direction Nationale de l'Alphabétisation n.d., b).

Besides the above, smaller-scale nonformal education is also organized by the Benin government in a number of other Gbe varieties, such as Saxwe and Waci in the Mono region and Tofin in the Atlantique region in southeastern Benin (Direction Nationale de l'Alphabétisation n.d., a,b,c,d). In Waci, literacy classes are conducted using the above-mentioned Waci-Gen primer.

2.3 PREVIOUS LINGUISTIC RESEARCH. The Gbe language varieties have been the focus of much linguistic research, especially during the last twenty years. Among others, such research was conducted by the national centers for applied linguistics and the national universities of Benin, Ghana, Nigeria, and Togo, as well as by American and European universities. Of special interest for the purposes of this paper are Capo's (1986, 1991) and Stewart's (1994) diachronic genetic studies.

Capo's (1986) comparative study of Gbe, which he began in 1971 and which was published under the title *Renaissance du gbe*, focuses on the phonological and morphophonological characteristics of the Gbe varieties. In a more recent publication, *A comparative phonology of Gbe*, Capo (1991:15) discusses his reconstruction of the consonant and vowel systems of Proto-Gbe from a sample of 19 Gbe varieties. Based on four innovation sets, he arrives at the internal classification of Gbe into the above-mentioned five branches which are, according to their geographical location from west to east, Ewe, Gen, Aja, Phla-Phera, and Fon (see Table 1; see also Figure 1 in section 2.1).

Ewe cluster		Gen cluster	Aja cluster	Phla-Phera cluster	Fon cluster
Adan	Kpelen	Agoi	Dogbo	Alada	Agbome
Agu	Peki	Anexo	Hwe	Ayizo	Kpase
Aveno	Towun	Gliji	Sikpi	Kotafon	Gun
Awlan	Ve		Tado	Saxwe	Maxi
Dayin	Vhlin			Tofin	Weme
Fodome	Vo			Toli	
Gbin	Waci			Xwela	
Ho	Wance			Xwla	
Kpando					

TABLE 1: Classification of Gbe varieties (Capo 1991:15)

In his review of Capo 1991, in questioning one of Capo's four sets of phonological innovations, Stewart (1994) arrives at an alternative internal classification of Gbe into two major clusters, i.e., Eastern and Western Gbe, with a further subdivision of Western Gbe into Ewe and Central Gbe (see Table 2).

Western Gbe		Eastern Gbe
Ewe	Central Gbe • Gen • Aja	Fon Phla-Phera

TABLE 2: Stewart's (1994:178) alternative internal classification of Gbe

3. SIL'S SYNCHRONIC STUDY OF GBE. In the context of SIL's larger extensibility study on Gbe, word and phrase lists were elicited in 49 Gbe varieties, chosen on the basis of Capo's (1986) comparative study of the Gbe language continuum.

The word and phrase lists were elicited from individual L1 speakers of the variety in question, not all of whom were residing in the language area. Ideally though, word and phrase lists should be elicited from groups of long-term resident L1 speakers. This procedure allows for the discussion of synonyms, near-synonyms, and variants, and group decisions can be made about which form to include in the lists, thus increasing reliability (see Sanders 1977:28f). Thus, the non-standard procedure applied during the study reported here potentially decreased the reliability of the elicited lists.

Given the overall goals of SIL's extensibility study, a synchronic rather than a diachronic approach was chosen for the analysis of the elicited linguistic data, with no attempt being made to establish the genetic relationships of the investigated Gbe varieties through a reconstruction of the main phonological features or principal grammatical traits.

The following sections briefly summarize the procedures and findings previously described in detail in Kluge (2005, 2006).

3.1 ANALYSIS OF LEXICAL SIMILARITY. The analysis of the 100-word list elicited in 49 Gbe varieties was conducted in four steps. Step one consisted of a qualitative analysis, and steps two to four consisted of a quantitative analysis.

The qualitative part of the analysis was conducted from a synchronic point of view applying the inspection method, which focuses on phonetic similarity to measure the "relative degrees of lexical relationship" of closely related languages as an indicator of potential intelligibility (Gudschinsky 1955). For further interpretation of word-list results in the context of extensibility studies, SIL's "Language assessment criteria" (International Language Assessment Conference 1990) recommend that if lexical similarity is more than about 70% (at the upper confidence limit of the calculation), dialect intelligibility testing should be conducted to explore the degree of intelligibility between the respective speech forms, whereas lexical similarity of less than 70% would suggest that the varieties involved are sufficiently different to be regarded as distinct languages.

Gudschinsky 1956 provides a set of guidelines for making lexical similarity decisions

that have since been adapted for use in linguistic and sociolinguistic language surveys (for example, by Blair 1990:30 ff.). However, while Gudschinsky's guidelines require a thorough morphemic analysis of the elicited lexical items to isolate equivalent morphemes, Blair's (1990) guidelines do not deal with the morphemic structure of the elicited lexical items but suggest that complete words be compared. SIL language survey teams have been applying Blair's guidelines to a large extent in their lexical analyses to making similarity decisions, since in most cases a thorough morphemic analysis proves rather unfeasible due to the limited scope of most SIL extensibility studies (see for example Dettweiler and Dettweiler 2003, Hochstetler et al. 2004, Meyers et al. 2003, Rueck and Jore 2003). The analysis findings are subsequently interpreted in terms of SIL's above-mentioned, recommended 70%-threshold level to establish priorities for further sociolinguistic research (see, for example, Brye and Brye 2004, Dettweiler and Dettweiler 2003, Harrison et al. 1999, and Hochstetler et al. 2004).

The 70% lexical similarity threshold, however, raises the question as to whether and to what extent the application of different similarity judgment criteria would result in differing computed lexical similarity degrees, and thereby in differing recommendations for further sociolinguistic research among the investigated Gbe varieties. Therefore, based on Probst's (1992) evaluation of different approaches in dealing with multimorphemic words in lexicostatistics, two different similarity judgment criteria sets were applied to the elicited Gbe data. Given the polymorphemic word structure to be found in a fair number of the elicited Gbe items, the first similarity judgment criteria set disregarded differences in the morphological structure. The second criteria set was more rigorous and did not disregard such differences; instead, Blair's (1990) guidelines were applied and pairs of complete words compared. To facilitate further analysis according to the two different similarity judgment criteria sets, the elicited data were organized in two different databases, employing WORDSURV (Wimbish 1989), a computer program designed for analyzing lexical similarity.

In the second step of the analysis, WORDSURV counted the lexical items identified as similar and computed two similarity percentage matrices according to the two different similarity judgment criteria sets. Given SIL's above-mentioned "Language assessment criteria" (International Language Assessment Conference 1990), the upper range of error was included in the computations.

The third step of the analysis involved paired t-tests and correlation analysis of both word-list computations, employing SPSS for Windows (SPSS Inc. 1998; see also Norušis 1993) and the package R (R-Project, n.d.). Paired t-tests were conducted to assess whether the different similarity judgment decisions resulted in significantly different lexical similarity percentages, whereas correlation analysis was employed to explore whether the two word-list computations indicate the same or different relative relationships between the investigated speech varieties.

The fourth step involved multidimensional scaling (MDS) of the computed similarity matrices, again employing SPSS for Windows, to investigate the relative relationships and clustering of the Gbe varieties, and thereby to verify the findings of the correlation analysis. Since the large number of varieties resulted in very dense configurations and rendered a more detailed interpretation difficult, three MDS plots each were computed for each computation: one including the complete set of 49 Gbe varieties, a second one including only

the western Gbe varieties, and a third one referring only to the eastern Gbe varieties. Employing a two-dimensional analysis, the computed plots were interpreted based on a combination of dimensional and neighborhood interpretation (see Kruskal and Wish 1978).

The results suggest that for the Gbe data set, the consistent application of different similarity judgment criteria sets yields compatible conclusions as to the relative relationships and clustering of the investigated varieties. More specifically, despite statistically significantly different lexical similarity percentages, the findings indicate for both computations the internal classification of the Gbe varieties into the same major Gbe clusters and subgroupings. Furthermore, the findings suggest that the less rigorous similarity judgment criteria, which disregard differences in the morphological structure, bring out similarities and differences less clearly, and thus yield possible alternative groupings that would, overall, affect the degree of priority assigned to the research of the respective varieties, but not the scope of such research.

3.2 ANALYSIS OF GRAMMATICAL SIMILARITY. Quantitative analysis of grammatical features focuses on the quantification of the syntactic and morphological systems of languages to explore “the frequency, distribution and relations of syntactic phenomena” and “of units, i.e., word-forms” (Těšitelová 1992:102, 119). (For two examples of such a quantitative approach, see Askedal 1977 and Greenberg 1960.)

Evidently uncommon is the quantitative analysis of grammatical features to obtain a rough estimate of the computed degrees of linguistic similarity among these language varieties, and thereby to identify the relative relationships and clustering of the investigated varieties. It appears that studies which investigate linguistic similarity across different speech varieties focus on the simple inspection of the elicited features without further quantification of the results (see for example Korhonen et al. 1986a, b and Sim 1980).

The only study, to my knowledge, that proposes the quantification of grammatical features to compute degrees of linguistic similarity among different speech varieties as a basis for identifying the clustering of these varieties is Wiesemann’s (1989) analysis of “Features of grammar in language assessment.” In the context of the previously mentioned accessibility issues pertinent to SIL’s extensibility studies, Wiesemann (1989) proposes the analysis of grammatical features in addition to lexicostatistical analysis to obtain a rough approximation of how the investigated speech varieties group and where intelligibility may be unlikely. Based on her research in a number of African languages across the continent, Wiesemann (1989:2) developed a French 35-phrase list that focuses on the noun and the verbal reference systems which, according to Wiesemann, are particularly pertinent to intelligibility (see also Wiesemann 1986 and Wiesemann et al. 1983:113 f.). During a preliminary study of 13 Gbe varieties, Wiesemann (1989) focused on the identification of the various strategies employed by the respective Gbe varieties to express relevant grammatical features, as well as judgments as to whether or not strategies for expressing a given grammatical feature are similar. Subsequently, Wiesemann (1989) conducted a quantification of these similarity judgments with the computer program LEXISTAT (Schadeberg 1990) which computed a dendrogram that indicated the clustering of the investigated Gbe varieties according to their quantified degrees of grammatical similarity.

Following this preliminary study, the complete Gbe data set consisting of 49 phrase lists was analyzed (see Kluge 2000, 2006 for a detailed discussion of the procedures of this

analysis). Again, the qualitative part of the analysis was conducted by Wiesemann (pers. comm., reported in Kluge 2000) and resulted in similarity judgments across the 49 Gbe varieties as to whether or not the strategies employed for expressing the various grammatical features are similar.

The second step of the analysis involved the quantification of these similarity judgments and the computation of a similarity matrix. Whereas Wiesemann (1989) had conducted this part of the analysis with LEXISTAT, I analyzed the data with WORDSURV. Based on Wiesemann's similarity judgments (pers. comm., reported in Kluge 2000), the elicited grammatical strategies were organized in a WORDSURV database and a similarity matrix was computed. However, given that WORDSURV was originally not designed to analyze grammatical features, and further given the rather low number of records for this computation, the calculations for grammatical similarity were not meant to imply the actual degrees of grammatical similarity but to indicate how the Gbe varieties are distributed into groups.

In the third step, as was the case for the synchronic analysis of lexical similarity, the computed phrase-list similarity matrix was analyzed by means of multidimensional scaling to assess the relative relationships and clustering of the Gbe varieties; again, three MDS plots were computed.

The results of the synchronic grammatical analysis indicate the internal classification of the Gbe varieties into the same three major Gbe clusters and subgroupings as indicated for the lexical similarity analysis. The indicated groupings were also compared to the findings of more recent Rapid Appraisal sociolinguistic surveys conducted among some of the Gbe communities and—although based on a synchronic typological rather than diachronic genetic approach—to the findings of Capo's (1991) and Stewart's (1994) diachronic genetic analyses. The findings of these comparisons suggest that, overall, the three different approaches result in similar classifications of the investigated Gbe varieties. However, there are also some significant differences regarding the classification of some of the varieties involved.

4. RESULTS. In the following sections, the major findings from the synchronic analysis of the elicited lexical and grammatical Gbe features are presented, focusing on the relative relationships and clustering of the Gbe varieties as indicated by the computed MDS plots. (For a more detailed presentation of the analysis findings, see Kluge 2005, 2006.)

With respect to the lexical similarity analysis, only one of the two above-mentioned matrices computed from word lists was chosen for the current analysis, namely the computation based on the more rigorous similarity judgment criteria set which does not disregard differences in the morphological structure. This criteria set applies the principles for lexical similarity decisions outlined by Blair (1990), principles which, as already mentioned, SIL language survey teams have been employing to a large extent in their lexical similarity comparisons. As deemed pertinent to the current analysis, though, the above-mentioned possible alternative groupings that were indicated for the computation based on the less rigorous similarity judgment criteria will also be discussed; these criteria disregard differences in the morphological structure.

An initial descriptive analysis of the computed similarity matrices suggests a chaining pattern for the entire Gbe cluster corresponding to the geographic location of the Gbe

speech varieties groups from Ghana's Volta region across southern Togo and Benin to the southwestern corner of Nigeria: overall, the investigated Gbe speech forms are marked by relatively small linguistic differences between adjoining speech varieties, whereas differences are greater between Gbe varieties at opposing ends of the chain.

4.1 STATISTICAL RELATIONSHIPS BETWEEN WLC AND PLC. The computation of the lexical and grammatical similarity decisions resulted in two different similarity matrices, WLC (word list computation) and PLC (phrase list computation). For the lexical data, the overall average degree of lexical similarity is 64%, with similarity percentage ranges of 45–100%, whereas for the grammatical data, the overall average degree of grammatical similarity is lower at 52% and the range larger with 21–100% ($n=1176$), thus yielding an observed difference of 12% between the mean WLC percentages and the mean PLC percentages.

Paired t-tests were conducted to assess whether the lexical and grammatical analyses resulted in significantly different similarity percentages. To obtain a set of approximately independent observations, 24 location pairs were chosen such that no location is in more than one pair. The results indicate that the observed mean difference between the WLC and PLC values for the 24 observations is 11.46%, which is statistically significantly different from 0 ($p\text{-value} < 0.0004$). Thus, it can be concluded that the two different similarity judgment criteria sets yield statistically significantly different similarity percentages.

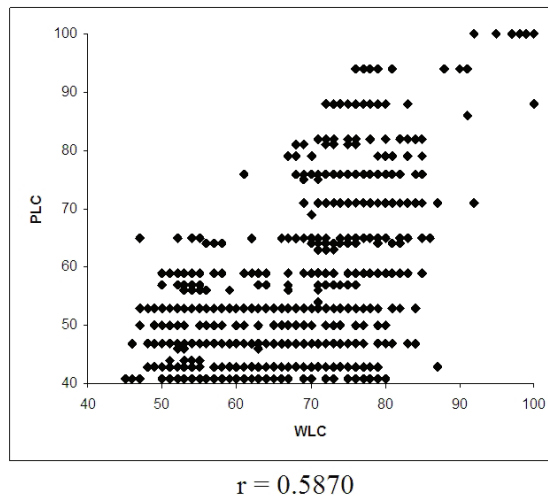


FIGURE 2: Scatterplot for word-list computations WLC and PLC—Complete data set

Correlation analysis of WLC and PLC results in a rather wide scatter, not necessarily indicating a linear and positive relationship (see Figure 2). This rather wide scatter leads to a correlation coefficient of $r = 0.5870$.

The correlation between WLC and PLC based on only the 24 nonoverlapping observations selected is $r = 0.5317$ (see Figure 3). An approximate 95% confidence interval for

the correlation between WLC and PLC is (0.1993, 0.7347) using the bootstrap method (Efron and Tibshirani 1993). Clearly, the correlation between the two similarity measures is not statistically significantly greater than 0.70. There is, however, statistical evidence for a positive correlation between WLC and PLC, although insufficient evidence to conclude that this correlation is very strong.

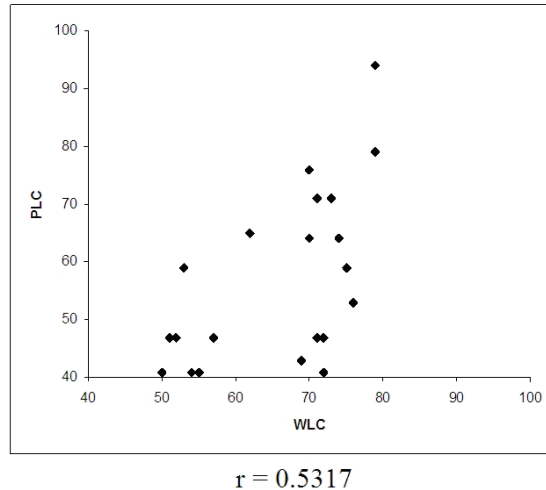


FIGURE 3: Scatterplot for word-list computations WLC and PLC—24 observations

4.2. GROUPING OF THE GBE LANGUAGE VARIETIES.

4.2.1 OVERVIEW. Multidimensional scaling of the computed word- and phrase-list matrices results in the MDS plots presented in Figure 4 and Figure 5, with Kruskal’s stress values of 0.246 and 0.264, and the squared simple correlation (RSQ) values of 0.857 and 0.71 for the word list computation (WLC) and the phrase-list computation (PLC), respectively, indicating that the derived configurations have an acceptable correlation with the input data.⁴ The ovals denote my interpretation of the clusters as suggested by dimensions 1 and 2 of the MDS plots. According to this interpretation, dimension 1 refers to the degree

⁴ MDS provides for each configuration two measures to test the results for reliability and validity: (1) a measure of stress (“Kruskal’s stress”) to measure how well the derived configuration matches the input data, and (2) the “squared simple correlation” (RSQ) to determine what proportion of variance of the scaled data can be accounted for by the MDS procedure. For Kruskal’s measure of stress, small values approaching 0.0 indicate an acceptable goodness of fit, whereas RSQ values of ≥ 0.60 are generally considered acceptable.

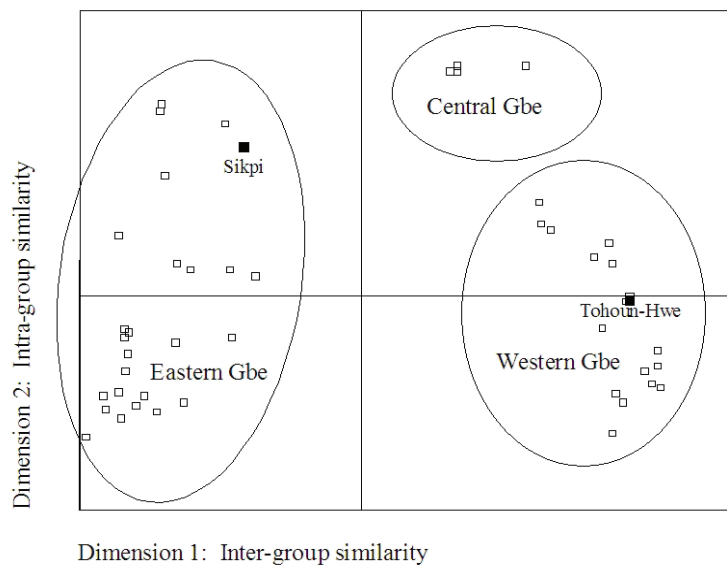


FIGURE 4: Major Gbe clusters – WLC-MDS plot

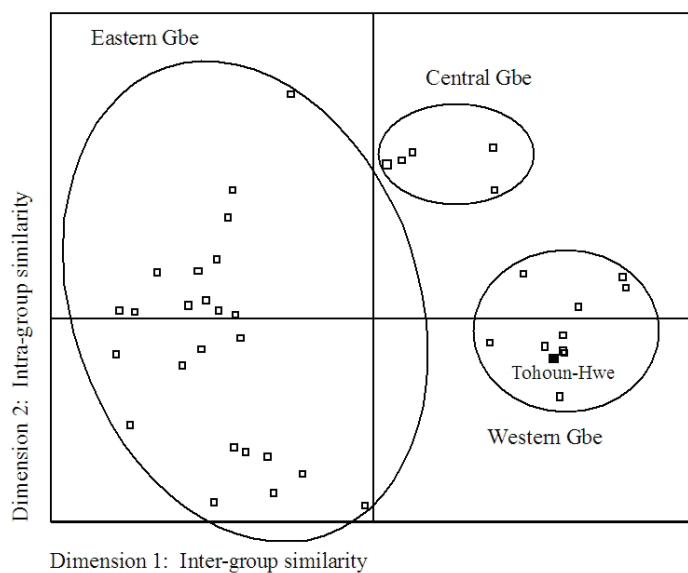


FIGURE 5: Major Gbe clusters—PLC-MDS plot

of intergroup similarity of the Gbe varieties, whereas dimension 2 refers to the degree of intragroup similarity for each cluster.

For both computations the MDS findings indicate three distinct clusters of Gbe varieties in terms of their degrees of intergroup similarity. Given their distinct geographical locations, these clusters are tentatively referred to as “Eastern, Central, and Western Gbe” (see Figure 1, section 2.1).

In this context, it is necessary to mention the following disparities between the MDS findings and the results suggested by a sociolinguistic survey conducted among two of the forty-nine investigated Gbe varieties, Tohoun-Hwe and Sikpi. Both MDS plots identify Tohoun-Hwe as a component of the western Gbe cluster. In addition, the WLC-MDS plot indicates Sikpi as a component of the eastern Gbe cluster, whereas the PLC-MDS plot suggests Sikpi as a component of the central Gbe cluster (see Figure 4 and Figure 5). In contrast, the findings of a more recent in-depth sociolinguistic survey of the Central Gbe, i.e., Aja communities, clearly identified Sikpi and Tohoun-Hwe as Aja varieties (Tompkins and Kluge 2002). Therefore, the results concerning Sikpi and Tohoun-Hwe are discussed together with the results for the remaining Aja varieties in “Central Gbe varieties” (section 4.2.3).

Table 3 presents an overview of the three identified major Gbe clusters and their components; given the above-mentioned findings of a more recent sociolinguistic survey of the Aja communities, Sikpi and Tohoun-Hwe are grouped together with the central Gbe varieties (for more details see section 4.2.3).

Western Gbe varieties		Central Gbe varieties	Eastern Gbe varieties		
Adan	Ho	Aja-Dogbo	Agbome	Gbesi	Seto
Agoi / Glijji	Kpelen	Aja-Hwe (Aplahoué)	Ajra	Gbokpa	Tofin
Agu	Kpesi	Aja-Hwe (Azovè)	Alada	Gun	Toli
Anexo	Togo	Aja-Hwe (Gboto)	Arohun	Kotafon	Weme
Aveno	Vhlin	Aja-Hwe (Tohoun)	Ayizo	Kpase	Xwela
Awlan	Vo	Aja-Sikpi	Ci	Maxi	Xwla (east.)
Be	Waci		Daxe	Movolo	Xwla (west.)
Gbin	Wance		Fon	Saxwe	
Gen	Wundi		Gbekon	Se	

TABLE 3: Major Gbe clusters with elicited Gbe varieties according to their larger geographical setting

In the following sections, detailed findings of multidimensional scaling of the computed word- and phrase-list matrices are presented according to the larger geographical clustering of the investigated varieties, i.e., the western, central, and eastern Gbe varieties.

4.2.2 WESTERN GBE VARIETIES. The western Gbe varieties are located in southeastern Ghana, southern Togo, and southwestern Benin, between the Volta and Dayi rivers to the west and the Mono river to the east (Capo 1986: map 1a) (see Figure 8 and Figure 9). The

findings of Capo's (1991:15) comparative phonology of Gbe indicate two clusters of western Gbe varieties, Ewe and Gen, each comprising the following speech varieties:

Ewe: Adan, Agu, Aveno, Awlan, Dayin, Fodome, Gbin, Ho, Kpando, Kpelen, Peki, Towun, Ve, Vhlin, Vo, Waci, and Wance

Gen: Agoi, Anexo, and Gliji

For the current analysis, lexical and grammatical data were elicited in the following western Gbe varieties:

Adan, Agoi/Gliji, Agu, Anexo, Aveno, Awlan, Be, Gbin, Gen, Ho, Kpelen, Kpesi, Togo, Vhlin, Vo, Waci, Wance, and Wundi

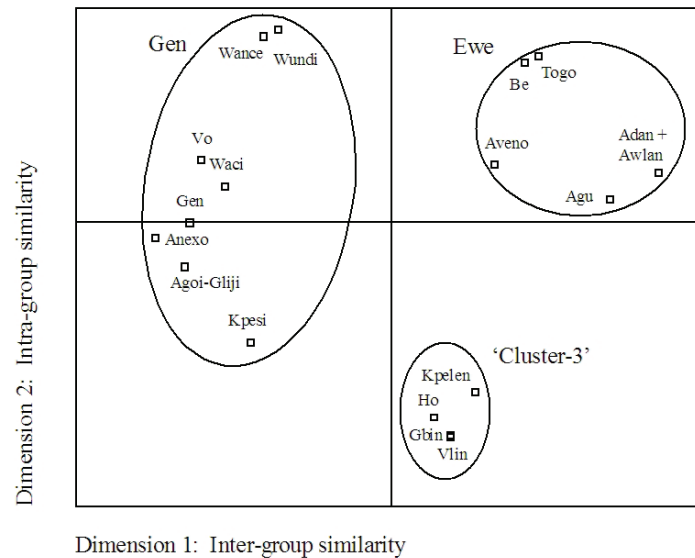


FIGURE 6: Western Gbe varieties – WLC-MDS plot

The results of multidimensional scaling of the computed word- and phrase-list matrices for the Western Gbe varieties are displayed in Figure 6 and Figure 7, with the MDS stress values of 0.256 and 0.165, and the RSQ values of 0.665 and 0.891, for WLC and PLC, respectively, suggesting that the derived configurations have an acceptable correlation with the input data. The ovals denote my interpretation of the clusters as indicated by dimensions 1 and 2 of the MDS plots. According to this interpretation, dimension 1 refers to the degree of intergroup similarity of the western Gbe varieties, whereas dimension 2 refers to the degree of intragroup similarity for each cluster.

The MDS findings suggest three distinct clusters of western Gbe varieties in terms of their degrees of intergroup similarity: the Gen and Ewe clusters and a distinct third cluster, preliminarily referred to as “northwestern Gbe.” This placement into three distinct clusters

appears to correspond to the more or less distinct geographical locations of these varieties, as indicated in Figure 8 and Figure 9.

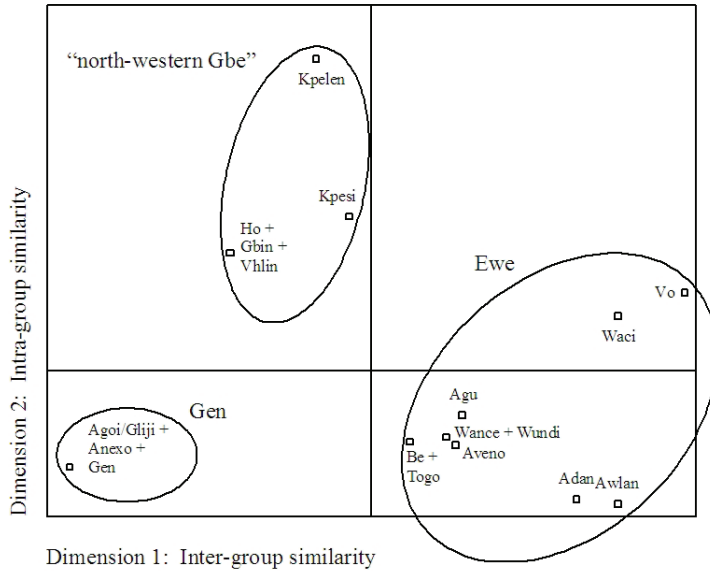


Figure 7: Western Gbe varieties – PLC-MDS plot

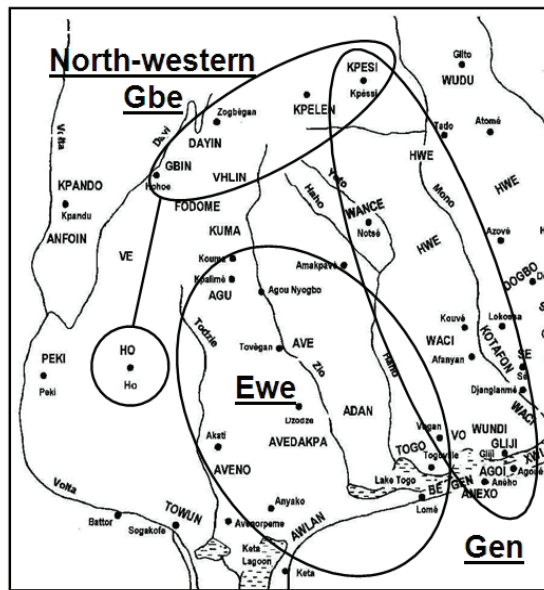


FIGURE 8: Geographical location of the Western Gbe varieties according to WLC-MDS

The Ewe cluster comprises the Adan and Awlan varieties, Awlan being the most prestigious variety of the Ewe cluster and the basis for Standard Ewe. The Agu, Aveno, Be, and Togo varieties are placed in the general neighborhood of Awlan and Adan, indicating the grouping of these varieties within the Ewe cluster. In addition, the PLC-MDS plot indicates Vo, Waci, Wance, and Wundi as components of the Ewe cluster.

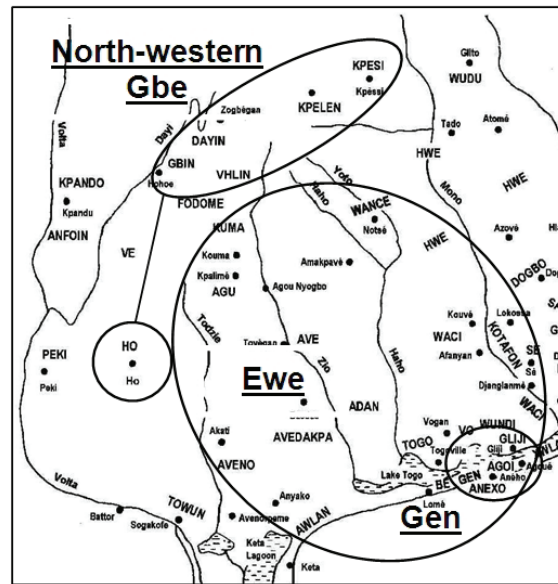


FIGURE 9: Geographical location of the Western Gbe varieties according to PLC-MDS

The Gen cluster comprises the Agoi/Gliji, Anexo, and Gen varieties, with the Gen variety already standardized and being used as the language of nonformal education in the southwestern part of Benin. In addition, the WLC-MDS plot indicates Kpesi, Vo, Waci, Wance, and Wundi as components of the Gen cluster.

The MDS analysis further indicates a distinct third cluster—preliminarily referred to as “northwestern Gbe” – which comprises Gbin, Ho, Kpelen, and Vhlin. In addition, the PLC-MDS plot suggests Kpesi as a component of the northwestern Gbe cluster. Kpesi, however, also shares a comparable degree of grammatical similarity with Vo and Waci, suggesting a possible alternative grouping of Kpesi with those varieties.

In terms of their degrees of intragroup similarity, the MDS findings indicate a rather compact placement of the northwestern Gbe varieties, the Ewe varieties (for WLC-MDS), and the Gen varieties (PLC-MDS), suggesting little intragroup variation. The placement of the Gen varieties (for WLC-MDS) and of the Ewe varieties (for PLC-MDS) is somewhat less dense, thus pointing to a higher degree of intragroup variation. These findings are evidenced by the respective ranges of intragroup lexical similarity, as discussed in more detail in Kluge 2005.

It should be noted though, that for the word-list computation based on the less rigorous similarity judgment criteria, which disregard differences in the morphological structure,

the MDS findings suggest possible alternative groupings for some of the western Gbe varieties. For the Agu, Aveno, Be, and Togo varieties, the findings suggest possible groupings with (1) the Gen cluster comprising the Agoi/Gliji, Anexo, Gen, Kpesi, Vo, Wance, Waci, and Wundi varieties, (2) the Adan and Awlan varieties, or (3) the Vo, Wance, Waci, and Wundi varieties in a distinct fourth cluster. Likewise, the findings indicate possible groupings for the Vo, Wance, Waci, and Wundi varieties with (1) the Agu, Aveno, Be, and Togo varieties in a distinct fourth cluster, or (2) the Agoi/Gliji, Anexo, Gen, and Kpesi varieties. (For more details see Kluge 2005.)

4.2.3 CENTRAL GBE VARIETIES. The grouping of central Gbe varieties is constituted entirely of the Aja varieties which are located in both Benin and Togo over a large area on both sides of the Mono river. In Benin, the Aja varieties are situated mainly in Benin's Mono region, in an area bordered by the Mono river to the west and the Kouffo river to the east. In Togo, the Aja varieties are situated west of the Mono river in the eastern part of the Maritime region and the southeastern corner of the Plateaux region (see Figure 10) (Tompkins and Kluge 2002).

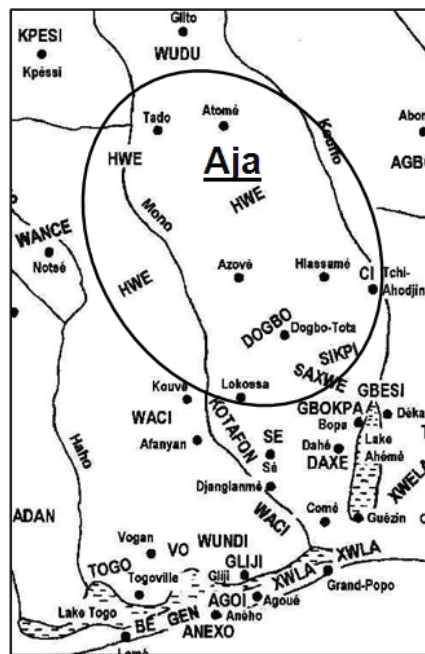


FIGURE 10: Geographical location of the Central (Aja) Gbe varieties

As previously mentioned, the Aja cluster was selected by the Benin government in 1992 for large-scale nonformal adult education (Direction de l'Alphabétisation 1992, see section 2.2).

According to Capó (1991:15), the Aja cluster comprises the following varieties: Dogbo, Hwe, Sikpi, and Tado. Within the context of the SIL study, word and phrase lists were elicited in the Dogbo variety, in the Hwe varieties spoken at Aplahoué, Azovè, Gboto, and Tohoun, and in the Sikpi variety.

The MDS findings for the computed word- and phrase-list similarity matrices indicate Dogbo, Aplahoué-, Azovè-, and Gboto-Hwe as components of Central Gbe, i.e., the Aja cluster. Neither Tohoun-Hwe nor Sikpi, though, are identified as components of the Aja language. Instead, the WLC- and PLC-MDS plots identify Tohoun-Hwe as a component of the western Gbe cluster, and the WLC-MDS plot indicates Sikpi as a component of the eastern Gbe cluster. In 1996, however, a more recent and in-depth sociolinguistic study of the Aja communities was conducted that included the elicitation of word lists, administration of sociolinguistic questionnaires, and comprehension testing. The findings of this study clearly indicate that Tohoun-Hwe and Sikpi are, in fact, varieties of Aja and, therefore, components of the central Gbe cluster (see Tompkins and Kluge 2002).

The disparities between the survey and the MDS findings could be because the word and phrase lists which formed the basis for the MDS analysis were not elicited from groups of long-term resident L1 speakers but from individual L1 speakers of the varieties in question, some of whom were not residing in the language area. Thus, the procedure did not allow for the discussion of variants, with group decisions made by L1 speakers resident in the area, as to which form to include in the lists, thus potentially decreasing reliability.

4.2.4 EASTERN GBE VARIETIES. The eastern Gbe varieties are located in Benin and Nigeria between the Mono river in southwestern Benin and the Yewa river in southwestern Nigeria, with the exception of a few western Xwla communities located to the west of the Mono river in southeastern Togo, in and around the town of Adamé (see Henson and Kluge 1999).

The results of Capó's (1991:15) diachronic study indicate two clusters of eastern Gbe varieties, Fon and Phla-Phera, each comprising the following speech varieties:

Fon: Agbome, Gun, Kpase, Maxi, and Weme

Phla-Phera: Alada, Ayizo, Kotafon, Saxwe, Tofin, Toli, Xwela, and Xwla

For the current analysis, lexical and grammatical data were elicited in the following eastern Gbe varieties:

Ajra, Agbome, Alada, Arohun, Ayizo, Ci, Daxe, Fon, Gbekon, Gbesi, Gbokpa, Gun, Kotafon, Kpase, Maxi, Movolo, Saxwe, Se, Seto, Tofin, Toli, Weme, Xwela, and eastern and western Xwla

Multidimensional scaling of the computed word- and phrase-list matrices for the Eastern Gbe varieties results in the MDS plots displayed in Figure 11 and Figure 12, with the MDS stress values of 0.284 and 0.28 and the RSQ values of 0.677 and 0.711 for WLC and PLC, respectively, indicating that the derived configurations have an acceptable cor-

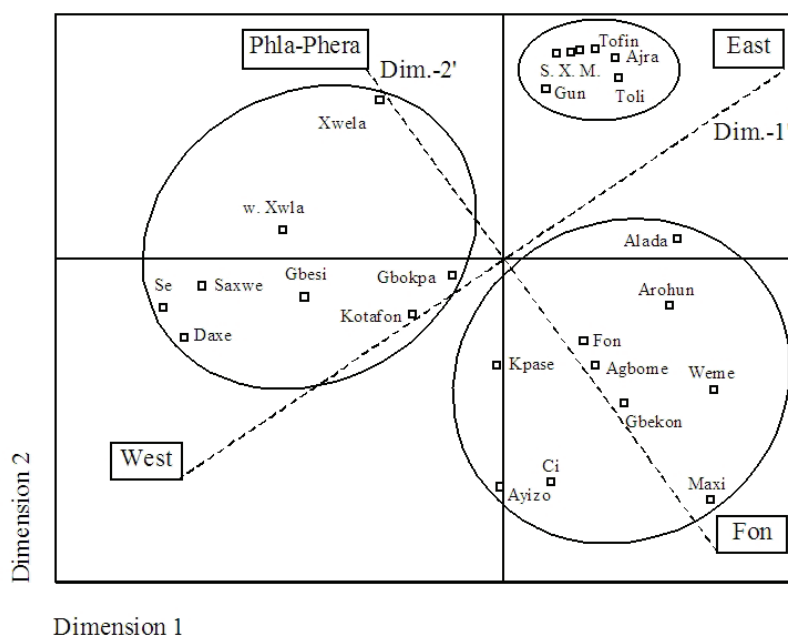


FIGURE 11: Eastern Gbe varieties – WLC-MDS plot (S.X.M. = Seto, eastern Xwla, Movolo)

relation with the input data. With regard to the dimensional interpretation of the displayed plots, it is noted though that dimension 1 and dimension 2 do not lend themselves to a clear interpretation. However, the MDS plots suggest a dimensional interpretation along the superimposed dashed lines. Thus, according to my interpretation, dimension 1' roughly suggests a perceptual east-west axis with west located to the left and east to the right, whereas dimension 2' indicates the extent to which the eastern Gbe varieties are grouped within the Fon or the Phla-Phera clusters.

Along the perceptual Fon – Phla-Phera axis both MDS plots suggest three distinct clusters: a Fon cluster situated on the Fon side of this axis, and a western and an eastern Phla-Phera cluster situated at its Phla-Phera side. Along the perceptual east-west axis, the three clusters are situated according to their actual geographical distribution (as indicated in Figure 13): the western Phla-Phera varieties are placed on the western side, the eastern Phla-Phera varieties on the eastern side, and the Fon cluster varieties in the center.

The Fon cluster comprises Agbome, Alada, Ayizo, Kpase, Maxi, and Weme, as well as the separate entry for Fon and the still unclassified Arohun, Ci, and Gbekon varieties. In addition, the PLC-MDS plot indicates Gun, Kotafon, and the still unclassified Gbesi and Gbokpa varieties as components of the Fon cluster, whereas the WLC-MDS plot suggests the grouping of these varieties with the Phla-Phera varieties.

The Phla-Phera varieties refer to Saxwe, Tofin, Toli, eastern and western Xlwa, and Xwela, as well as the still unclassified Ajra, Daxe, Movolo, Se, and Seto varieties. In addition, as already mentioned, the WLC-MDS plot indicates Gbesi, Gbokpa, Gun, and Kota-

fon as Phla-Phera varieties. For Gbokpa and Kotafon, however, the respective WLC lexical similarity percentages also indicate a possible alternative grouping within the Fon cluster.

The MDS results further indicate two distinct groupings of Phla-Phera varieties, a western and an eastern one, which correspond to the actual geographical distribution of these varieties (as indicated in Figure 13). Moving roughly from east to west, the indicated eastern Phla-Phera cluster comprises:⁵ Seto, Ajra, Toli, eastern Xwla, Movolo, (Gun), and Tofin, all of which are located in the southeastern corner of Benin, close to the Nigerian border, whereas the indicated western Phla-Phera cluster comprises: (Gbesi, Gbokpa),

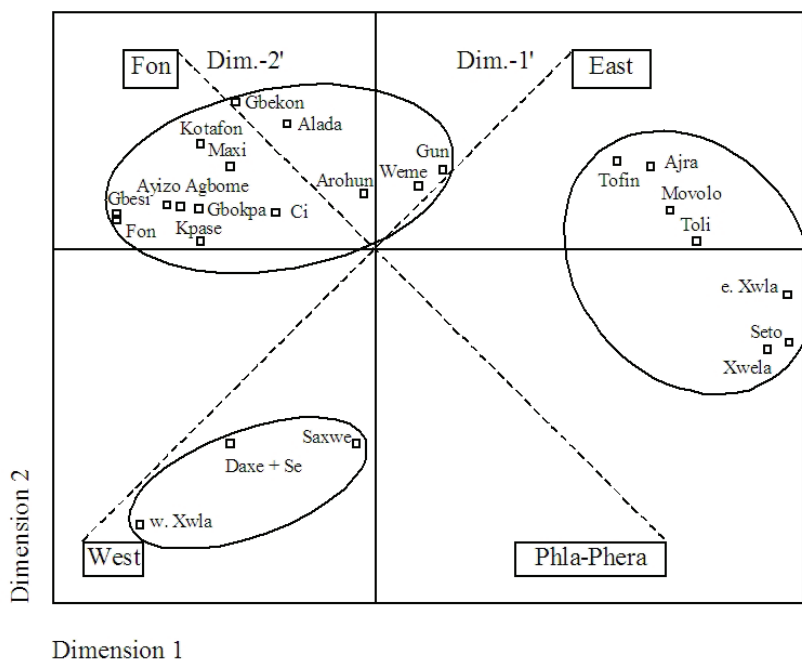


FIGURE 12: Eastern Gbe varieties – PLC-MDS plot

Daxe, Saxwe, Se, (Kotafon), and western Xwla, all of which are located in the southwestern corner of Benin, close to the Togolese border. For Xwela, the findings are ambiguous, with the WLC-MDS plot indicating Xwela as a western Phla-Phera variety and the PLC-MDS plot indicating it as an eastern Phla-Phera variety.

With the exception of Kotafon, Toli, and Xwela, this grouping of the eastern Gbe varieties into three distinct clusters appears to correspond to the more or less distinct geographical locations of these varieties.

The Kotafon variety, which the PLC-MDS findings indicate as a component of the Fon cluster, is located in the western part of the western Phla-Phera area, and thus is not in the

⁵ Eastern Gbe varieties that are indicated as Phla-Phera varieties by only one of the two MDS plots are listed in parenthesis.

immediate neighborhood of other Fon-speaking communities. The Toli area, indicated as an eastern Phla-Phera cluster component, is located in the central part of southern Benin, in and around Tori-Bossito, according to Capo (1986: map 1a), and thus is geographically closer to the western rather than the eastern Phla-Phera cluster varieties. The findings of a more recent Rapid Appraisal sociolinguistic survey, however, indicate that Toli of Tori-Bossito refers to one dialect of the Ayizo variety, namely “Tori” (Hatfield and McHenry 1998). A distinct Toli variety does exist though and is located in the southeastern corner of Benin, and thus is close to the eastern Phla-Phera cluster varieties; it was here that the Toli word- and phrase lists were elicited. Therefore, the map in Figure 13 includes an additional entry for Toli situated in the eastern Phla-Phera language area. The Xwela area, while indicated by the grammatical analysis findings as an eastern Phla-Phera cluster component, is situated on the eastern fringes of the western Phla-Phera cluster area, and thus is close to the western Phla-Phera cluster varieties.

5. DISCUSSION AND CONCLUSIONS. In this discussion, special attention is given to the question as to whether and to what extent the findings of the lexical and the grammatical

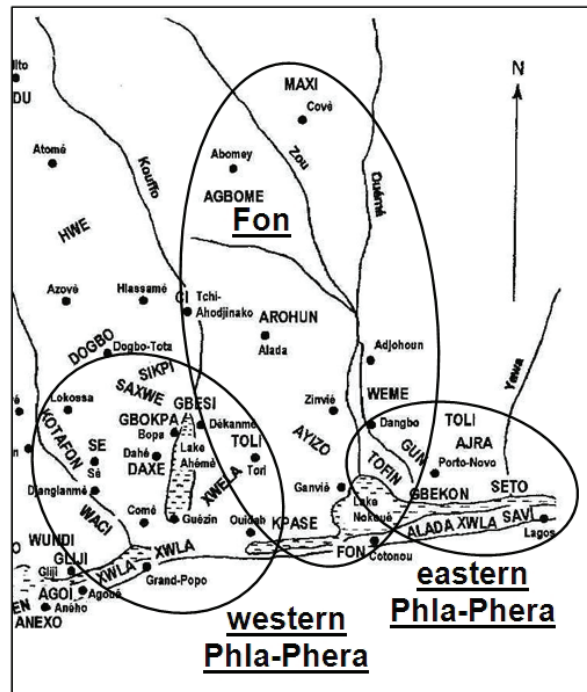


FIGURE 13: Geographical location of the Central and Eastern Gbe varieties

analyses yield the same or different conclusions regarding the relative relationships and clustering of the investigated varieties. Furthermore, this discussion explores how these

findings compare to the findings of more recent Rapid Appraisal sociolinguistic surveys, conducted among some of the Gbe communities, and—although based on a synchronic rather than diachronic approach—to the findings of Capo’s (1991) and Stewart’s (1994) genetic studies. This comparison is followed by a discussion as to what weight should be given to the different approaches and what conclusions can be drawn in terms of prioritizing focused sociolinguistic research in the context of an extensibility study of a language continuum.

5.1 CLUSTERING OF THE GBE LANGUAGE VARIETIES. In this section, the findings from the rather different approaches—that is, lexical versus grammatical similarity, synchronic versus diachronic, and linguistic versus sociolinguistic—are compared. These comparisons are not based on the assumption that for the synchronic approach the computed lexical and grammatical similarity percentages *per se* are comparable. Nor is it implied that the results from the synchronic analysis would indicate the genetic relationships of the investigated language varieties and therefore would be equivalent to the findings of the diachronic studies conducted by Capo (1991) and Stewart (1994). Rather, the focus of this comparison is the relative relationships and clustering of the investigated Gbe varieties as derived from multidimensional scaling of the computed similarity matrices.

The qualitative and quantitative analysis of lexical and grammatical features elicited in 49 Gbe varieties resulted in the computation of word-list and phrase-list similarity matrices. The findings from the paired t-tests indicated that for both matrices, the computed similarity percentages differ significantly. At the same time, though, the results of the correlation analysis suggest a positive correlation between both computations, although there is insufficient evidence to conclude that this correlation is very strong.

The findings of the MDS analysis of the word-list and the phrase-list computations enhance the conclusion of a positive correlation between both computations by indicating the same three major Gbe clusters: a western, a central, and an eastern one. In terms of the grouping of the 49 Gbe varieties into the three major clusters, the findings indicate 100% agreement between the findings of the lexical and the grammatical similarity analyses (see Table 4).

Furthermore, the findings indicate the following subgroupings: western Gbe comprises the Ewe and Gen clusters, and a distinct third cluster, tentatively referred to as “north-western Gbe cluster,” central Gbe consists entirely of the Aja varieties, and eastern Gbe comprises the Fon, western, and eastern Phla-Phera clusters (see Figure 14).

The findings of the synchronic analyses agree to a large extent with the findings of Capo’s (1991) and Stewart’s (1994) comparative studies, but there are also some significant differences.

Corresponding to the results of Stewart’s (1994) work, the findings of the synchronic analysis indicate three large groupings of Gbe varieties, i.e., a western, central, and eastern Gbe cluster, although, in contrast to Stewart (1994), the Gen varieties are grouped among the western Gbe varieties. While Capo’s (1991) classification does not suggest three large groupings such as a western, central, and eastern Gbe cluster, it is noted that all of Capo’s Ewe and Gen cluster varieties (13/13) are included in the suggested Western Gbe cluster. Likewise, Capo’s Aja varieties (3/3) are included in the indicated Central Gbe cluster, and all of Capo’s Fon and Phla-Phera cluster varieties (13/13) are included in the suggested

		Western Gbe		Central Gbe / Aja		Eastern Gbe			
		WLC	PLC	Capo	WLC/PLC	Capo	WLC	PLC	Capo
Ewe		Adan	Adan	Adan	Dogbo	Dogbo	Ajra	Ajra	
		Awlan	Awlan	Awlan	Hwe-Aplahoué	Hwe	Movolo	Movolo	
		Agu*	Agu	Agu	Hwe-Azove		Seto	Seto	
		Aveno*	Aveno	Aveno	Hwe-Gboto		Tofin	Tofin	Tofin
Gen		Be*	Be	Be	Hwe-Tohoun		Toli	Toli	Toli
		Togo*	Togo	Togo	Sikpl	Sikpl	Xwla-east	Xwla-east	Xwla
		Wance*	Wance	Wance			Gun	Gun	Gun
		Wundr*	Wundi	Wundi			Agbome	Agbome	Agbome
North-W. Gbe		Waci*	Waci	Waci			Arohun	Arohun	
		Vo*	Vo	Vo			Ci	Ci	
		Kpesi	Kpesi*	Kpesi*			Fon	Fon	
		Kpelen	Kpelen	Kpelen			Kpase	Kpase	Kpase
Gen		Gbin	Gbin	Gbin			Maxi	Maxi	Maxi
		Ho	Ho	Ho			Weme	Weme	Weme
		Vhlin	Vhlin	Vhlin			Gbekon	Gbekon	
		Agoi/Giji	Agoi/Giji	Agoi/Giji			Alada	Alada	Alada
	Anexo	Anexo	Anexo			Ayizo	Ayizo	Ayizo	
	Gen	Gen	Gen			Gbokpa#	Gbokpa		
						Kotafon#	Kotafon	Kotafon	
						Gbesi	Gbesi		
						Xwela	Xwela	Xwela	
						Daxe	Daxe		
						Saxwe	Saxwe	Saxwe	
						Se	Se		
						Xwla-west	Xwla-west	Xwla	

[Gbe varieties are italicized for which the findings indicate possible alternative groupings; * indicates a possible alternative grouping with Gen, + indicates a possible alternative grouping with Ewe, and # indicates possible alternative grouping with Fon]

TABLE 4: Clustering of the Gbe varieties according to the findings of the synchronic lexical and grammatical analyses and Capo's (1991) diachronic study

Eastern Gbe cluster. However, in contrast to Capo (1991), the findings of the current analyses suggest (a) a distinct third cluster of western Gbe varieties, preliminarily referred to as “northwestern Gbe,” that comprises western Gbe varieties that Capo (1991) classifies as components of the Ewe cluster, and (b) the grouping of the Phla-Phera varieties into two distinct clusters, western and eastern Phla-Phera.

Regarding the clustering of the 49 Gbe varieties within the suggested subgroupings, the findings suggest, overall, high levels of agreement between the lexical and the grammatical analyses: 40/49 Gbe varieties (82%) are indicated as components of the same subgroupings, i.e. 13/18 western varieties (72%), 6/6 central varieties (100%), and 21/25 eastern varieties (84%) (see Table 5 a and b).

The discrepancies between both analyses refer to 5 out of 18 western Gbe varieties (28%), namely Kpesi, Vo, Waci, Wance, and Wundi, and 4 out of 25 eastern Gbe varieties (16%), namely Gbesi, Gbokpa, Gun, and Kotafon. The possible alternative grouping of Kpesi within the Ewe cluster, as indicated by the grammatical analysis findings, would not affect the agreement levels, whereas the possible alternative grouping of two eastern Gbe varieties, namely Gbokpa and Kotafon, within the Fon cluster, as indicated by the lexical analysis findings, would increase the levels of agreement from 21 out of 25 varieties (84%) to 23 out of 25 varieties (92%).

In contrast, the above-mentioned possible alternative grouping for some of the western Gbe varieties, suggested by the MDS findings for the word-list computation based on the less rigorous similarity judgment criteria that disregard differences in the morphological structure, would lower the agreement levels. If the Agu, Aveno, Be, and Togo varieties were to be grouped together with the Vo, Wance, Waci, and Wundi varieties either within the Gen cluster or in a distinct fourth cluster, the agreement levels would decrease from 13 out of 18 varieties (72%) to 9 out of 18 varieties (50%).

For five of the investigated varieties, namely Kpesi, Ayizo, Kotafon, Xwela, and Tohoun-Hwe, the findings of more recent Rapid Appraisal sociolinguistic surveys indicate different or alternative outcomes as to their classification when compared to the findings of the synchronic analyses.

a) Kpesi is indicated as a Gen cluster variety, according to the lexical analysis findings. In contrast the grammatical analysis findings suggest Kpesi as a component of “northwestern Gbe,” while at the same time suggesting a possible alternative grouping of Kpesi within the Ewe cluster. The results of a more recent Rapid Appraisal sociolinguistic survey of the Kpesi communities modify these findings and support the grouping of Kpesi within the Ewe cluster (Durieux-Boon et al. 1998:9).

b) Ayizo is suggested to be a component of the Fon cluster, according to the findings of both synchronic analyses. Kotafon is indicated as a Fon cluster variety, according to the grammatical analysis findings, and as a Phla-Phera cluster variety, according to the lexical analysis findings. In contrast, the findings of more recent Rapid Appraisal surveys support the grouping of both varieties together with the Phla-Phera varieties, at least as far as the eastern Phla-Phera varieties Tofin and Toli are concerned, whereas regarding Xwela and Xwla this

grouping would appear to be more questionable (Hatfield and McHenry 1998; Hatfield et al. 1998).

c) Xwela is indicated by the lexical analysis findings as a western Phla-Phera variety and by the grammatical analysis results as an eastern variety. These findings are modified by the results of a more recent Rapid Appraisal sociolinguistic survey conducted among the Xwela communities that support the grouping of Xwela with the western Phla-Phera varieties (Henson et al. 1999).

d) For a discussion of Tohoun-Hwe, see “Central Gbe varieties” (section 4.2.3).

Levels of agreement according to analysis approach

	Western Gbe	Central Gbe	Eastern Gbe	Total
WLC – PLC	13/18 – 72%	6/6 – 100%	21/25 – 84%	40/49 – 82%
WLC – Capo	6/13 – 46%	3/3 – 100%	10/13 – 77%	19/29 – 66%
PLC – Capo	9/13 – 69%	3/3 – 100%	10/13 – 77%	22/29 – 76%

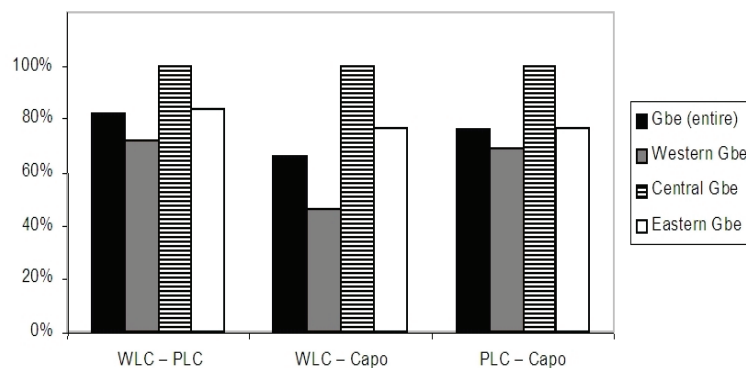


TABLE 5 (A): Levels of agreement between the synchronic lexical and grammatical analyses and Capo's (1991) diachronic analysis

The levels of agreement between the findings of the synchronic analyses and Capo's (1991) diachronic study are somewhat lower. Of the 29 Gbe varieties mentioned by Capo (1991), the lexical analysis findings indicate 19 varieties (66%) as components of the same subgroupings, i.e. 6 out of 13 western Gbe varieties (46%; counting only one entry for the Agoi and Gliji varieties), 3 out of 3 central varieties (100%), and 10 out of 13 eastern varieties (83%; counting Xlwa only once), while for the grammatical analysis findings the levels of agreement are slightly higher with 22 varieties (76%) being suggested as components of the same subgroupings, i.e. 9 out of 13 western Gbe varieties (67%), 3 out of 3 central

Levels of agreement according to the major Gbe clusters

	Western Gbe	Central Gbe	Eastern Gbe	Total
WLC – PLC	13/18 – 72%	6/6 – 100%	21/25 – 84%	40/49 – 82%
WLC – Capo	6/13 – 46%	3/3 – 100%	10/13 – 77%	19/29 – 66%
PLC – Capo	9/13 – 69%	3/3 – 100%	10/13 – 77%	22/29 – 76%

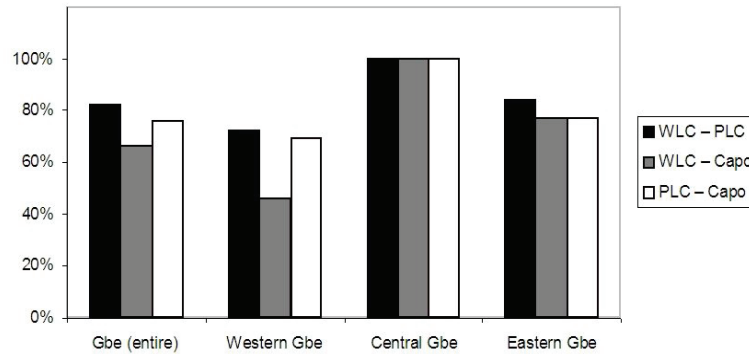


TABLE 5(B) : Levels of agreement between the synchronic lexical and grammatical analyses and Capo's (1991) diachronic analysis

varieties (100%), and 10 out of 13 eastern varieties (78%) (see Table 5 a and b).

For the synchronic lexical approach, the discrepancies to Capo's (1991) classification refer to 10 out of 29 varieties (33%), that is 7 out of 13 western Gbe varieties (54%) and 3 out of 13 eastern Gbe varieties (23%), whereas for the synchronic grammatical approach, the differences refer to 7 out of 29 varieties (24%), that is, 4 out of 13 western Gbe varieties (31%) and 3 out of 13 eastern Gbe varieties (23%).

More specifically, for the western Gbe varieties, the findings of both synchronic analyses suggest the grouping of the Gbin, Ho, Kpelen, and Vhlin varieties within in a distinct third cluster, which concurs with the distinct geographical location the farthest north of three of the four varieties, namely Gbin, Kpelen, and Vhlin; Ho is located farther south. In contrast, Capo (1991) indicates all four varieties as components of the Ewe cluster. In addition, the lexical analysis findings also indicate the Vo, Waci, and Wance varieties as components of the Gen cluster varieties, whereas Capo (1991) lists these varieties among the Ewe cluster varieties. Regarding Waci, it is notable that Waci literacy classes are conducted with the above-mentioned Waci-Gen primer that employs vocabulary shared by Gen and Waci, a procedure that suggests that both varieties are fairly closely related (see section 4.2.2) (Capo 1986:44f. ; 102, Direction Nationale de l'Alphabétisation n.d., b).

The possible alternative grouping of Kpesi within the Ewe cluster, as indicated by the grammatical analysis findings, would have no effect on the agreement levels. In contrast, the above-mentioned possible alternative grouping for some of the western Gbe varieties, suggested by the MDS findings for the word-list computation based on the less rigorous similarity judgment criteria, would lower the agreement levels between the synchronic

lexical analysis and Capo's (1991) diachronic study: if the Agu, Aveno, Be, and Togo varieties were to be grouped together with the Vo, Wance, Waci, and Wundi varieties, either within the Gen cluster or in a distinct fourth cluster, the agreement levels would decrease considerably from 6 out of 13 identical components (46%) to 2 out of 13 varieties (15%).

For the eastern Gbe varieties, discrepancies between the synchronic and diachronic approaches refer to 3 out of 13 varieties, two of which are Alada and Ayizo, which the findings of the synchronic approach suggest as components of the Fon cluster, whereas Capo (1991) lists them among the Phla-Phera varieties. In addition, the grammatical analysis findings suggest Kotafon as a Fon cluster variety, whereas Capo (1991) lists this variety also with the Phla-Phera varieties. Furthermore, the lexical analysis findings indicate Gun as a Phla-Phera variety, while Capo (1991) classifies Gun as a component of the Fon cluster. The possible alternative grouping of Kotafon within the Fon cluster, as indicated by the lexical analysis findings, would decrease the levels of agreement with Capo's (1991) classification from 10 out of 13 identical components (77%) to 9 out of 13 (70%).

5.2 WEIGHING DISPARITIES: A MULTIFACETED APPROACH TO LINGUISTIC DATA INTERPRETATION. The above comparisons indicate significant agreement between findings of the synchronic lexical and the synchronic grammatical analyses in terms of the relative relationships of the investigated Gbe varieties. The levels of agreement between the findings of the synchronic approaches and Capo's (1991) and Stewart's (1994) diachronic genetic studies are, overall, also fairly high. However, the comparisons also suggest some considerable disparities, and for the western Gbe varieties the levels of agreement between Capo's (1991) classification and the findings of the synchronic lexical analysis are only moderate. These disparities raise the question how the different approaches should be weighted and what conclusions can be drawn in terms of the ways of investigating and classifying dialects in a linguistic continuum.

SIL's synchronic analysis of the elicited linguistic features is, as already mentioned, embedded in SIL's larger sociolinguistic study of Gbe which has as its main objective the exploration of the extensibility of already existing materials among speakers of related speech varieties. Giving special attention to the previously mentioned accessibility and acceptability issues, this sociolinguistic approach focuses on the social uses of the investigated varieties in communication. To adequately inform language development program decisions in terms of literature extensibility, in-depth sociolinguistic studies would need to be conducted in all the varieties under investigation, an undertaking that would be rather unfeasible in the context of large language continua such as the Gbe continuum.

In drawing on the findings available from a variety of different approaches, for example, the synchronic lexical and grammatical analyses, Rapid Appraisal sociolinguistic surveys, as well as comparative studies, however, such in-depth sociolinguistic studies may not be necessary. Instead, it is proposed here to apply a multifaceted approach that considers the findings from the different methodologies as complementing each other, and thereby providing a better understanding of the overall language situation than would each of them taken individually. In this manner, different outcomes from the various studies would be used to inform focused in-depth sociolinguistic research investigating the pertinent accessibility and acceptability issues.

Regarding the question of how the different approaches should be weighted in light

of the identified disparities, it is proposed here that in the context of extensibility studies, more weight be given to the findings provided by in-depth sociolinguistic surveys. Such is the case with Tohou-Hwe and Sikpi, which the findings of an in-depth sociolinguistic survey of the Aja communities identified as a central Gbe, i.e., Aja varieties, while the synchronic linguistic analysis indicated Tohou-Hwe as a western and Sikpi as an eastern Gbe variety (Tompkins and Kluge 2002).

This study demonstrates how such a multifaceted approach could be undertaken. In comparing the findings of the synchronic lexical and grammatical analyses with each other and to the results of more recent sociolinguistic surveys, as well as to the findings of Capo's (1991) and Stewart's (1994) diachronic studies, it is suggested that the above-mentioned disparities not be regarded as contradictions, but be the focus of further research to:

1. Establish the grouping of the varieties in question more precisely

For the varieties in question, conduct focused in-depth sociolinguistic research to determine more precisely how these varieties should be classified in terms of the pertinent accessibility and acceptability issues. More specifically, for the western Gbe varieties, determine more precisely whether (1) the Gbin, Ho, Kpelen, and Vhlin varieties constitute a distinct cluster or group together with the Ewe varieties, (2) Kpesi is a component of such a distinct third cluster or a component of the Ewe cluster, (3) the Agu, Aveno, Be, and Togo varieties group within the Ewe cluster, or together with the Vo, Wance, Waci, and Wundi varieties either within the Gen cluster or within a distinct fourth cluster. All of these varieties are classified by Capo (1991) as components of the Ewe cluster. Interestingly, though, all of these varieties are situated on the fringes of the "Ewe" language area. Gbin, Kpelen, Kpesi, and Vhlin are the most northern Western Gbe varieties (Ho is located farther south), and the Waci, Wance, Wundi and Vo varieties are the most eastern Western Gbe varieties with Waci, Wundi and Vo also in close neighborhood to the Gen cluster varieties.

For the eastern Gbe varieties, establish whether (1) Alada, Ayizo, Gbesi, Gbokpa, Gun, and Kotafon are components of the Fon cluster or group together with the Phla-Phera varieties, (2) the Phla-Phera varieties group in two distinct clusters, i.e., western and eastern Phla-Phera, and (3) Xwela groups with the western or eastern Phla-Phera varieties.

2. Investigate extensibility of already existing materials to those varieties that are components of a cluster that has not yet undergone widespread language-based development

For the larger Gbe study, assess whether the "northwestern Gbe" communities could benefit from existing literacy efforts in Ewe or Gen, and whether the western and eastern Phla-Phera cluster communities could benefit from existing literacy efforts in Fon, Gen, or Gun, or whether additional language development programs in some of these clusters would be beneficial. In case additional development programs are deemed beneficial for the Phla-Phera communities, special emphasis should be given to the question of whether the eastern and western Phla-Phera varieties could benefit from the same additional development programs.

Given the chaining character of the Gbe varieties, the location of Gbin, Kpelen, Vhlin, and possibly Kpesi, on the fringes of Capo's (1991) Ewe area in itself raises the question

of whether and to what extent these varieties are linguistically close enough to the Awlan variety, upon which standard Ewe is based and which is situated in the very southern part of the “Ewe” area, to benefit from Ewe literacy efforts.

3. Investigate the underlying reasons for these disparities

Disparities between the synchronic lexical, the synchronic grammatical and the diachronic genetic approach could be due to (1) flawed elicitation procedures in that an inexperienced transcriber wrote down the data, (2) flawed elicitation procedures in that the word and phrase lists were elicited from individual rather than from groups of L1 speakers of the variety in question, (3) incompatibility of the lexical and grammatical structures investigated in the context of the synchronic analyses, (3) incompatibility of the synchronic and diachronic approaches, (4) drawbacks of the tools applied for the synchronic analyses, or (5) problems with the classification of Gbe varieties as suggested by the findings of Capo’s (1991) comparative analysis.

5.3 CONCLUSIONS. Taking as its point of departure a synchronic perspective, the study reported here involved a qualitative and quantitative analysis of lexical and grammatical features elicited among the Gbe language varieties of West Africa. Conducted in the context of the larger Gbe extensibility study, the main objective of this analysis was to explore how the investigated varieties might be treated as clusters and to establish priorities for further sociolinguistic research.

The comparison of the findings from the synchronic lexical and synchronic grammatical analyses, as well as the comparison of these findings to the outcomes of more recent sociolinguistic surveys and to the results of Capo’s (1991) and Stewart’s (1994) diachronic genetic studies indicate substantial agreement in terms of the relative relationships of the Gbe varieties investigated. However, there are also some significant differences regarding the classification of some of the varieties involved. Given the different purposes these various approaches are serving, though, it was not necessary to set these different approaches in opposition to each other. Rather, this paper has shown how in employing a multifaceted approach to linguistic data interpretation different, even contrasting, outcomes may be used to inform focused in-depth sociolinguistic research. Such research would in turn directly influence language program development decisions as to the extensibility of already existing literature to the remaining speech communities of a language continuum.

Dealing with but a single set of speech varieties investigated within the limited scope of an extensibility study, the approach offered here is suggested as a case study for examining the usefulness (1) of employing a synchronic typological perspective that focuses on the relative relationships and clustering of the investigated varieties by means of multidimensional scaling, and (2) of comparing, where available, the synchronic analysis findings to the outcomes of diachronic genetic studies and sociolinguistic surveys to inform focused in-depth sociolinguistic research. Comparable studies in different language situations, applying the methodology proposed here, are needed to substantiate the soundness of this approach chosen for the analysis of the Gbe data.

REFERENCES

- ASKEDAL, JOHN O. 1997. Computing morphosyntactic differences. A typological perspective on German and Norwegian. *Journal of Quantitative Linguistics* 4(1–3):35–44.
- BENNETT, PATRICK R., and JAN P. STERK. 1977. South Central Niger-Congo: A reclassification. *Studies in African Linguistics* 8:241–273.
- BERGMAN, TED G. 1991. Rapid appraisal of languages. *Notes on Sociolinguistics* 28:3–11.
- BLAIR, FRANK. 1990. *Survey on a shoestring: A manual for small-scale language surveys*. Dallas: Summer Institute of Linguistics and University of Texas.
- BRYE, EDWARD, and ELIZABETH BRYE. 2004. Intelligibility testing survey of Bebe and Kezemezung and synthesis of sociolinguistic research of the Eastern Beboid cluster. *SIL Electronic Survey Reports 2004-011*. Dallas: SIL International. <http://www.sil.org/silesr/2004/silesr2004-011.pdf>.
- CAPO, HOUNKPATI B.C. 1986. *Renaissance du gbe. Une langue de l'Afrique occidentale. Etude critique sur les langues ajatado: l'ewe, le fon, le gen, l'aja, le gun, etc.* Université du Bénin. Institut National des Sciences de l'Éducation. Etudes et Documents de Sciences Humaines. Série A: Etudes, Numéro 13. Lomé. Also published as: Capo, Hounkpati B.C. 1988. *Renaissance du gbe: Réflexions critiques et constructives sur l'ewe, le fon, le gen, l'aja, le gun, etc.* Hamburg: Helmut Buske.
- CAPO, HOUNKPATI B.C. 1991. *A comparative phonology of Gbe*. Publications in African Languages and Linguistics 14. Berlin: Walter de Gruyter.
- DETTWEILER, STEPHEN, and SONIA DETTWEILER. 2003. Sociolinguistic survey of the Duka (Hun-Saare) People. *SIL Electronic Survey Reports 2003-014*. Dallas: SIL International. <http://www.sil.org/silesr/2003/silesr2003-014.pdf>.
- DIRECTION DE L'ALPHABÉTISATION, RÉPUBLIQUE DU BÉNIN, MINISTÈRE DE LA CULTURE ET DES COMMUNICATIONS (DA). 1992. *Séminaire nationale sur la redéfinition des objectifs et stratégies d'alphabétisation et d'éducation des adultes. Rapport final*. Cotonou: Presse de Onepi.
- DURIEUX-BOON, EVELIN I. K., JUDE A. DURIEUX, DEBORAH H. HATFIELD, and BONNIE J. HENSON. 1998. A sociolinguistic survey of the Kpesi language area. Ms., SIL International, Cotonou.
- DUTHIE, ALAN S. 1988. Ewe. In *The languages of Ghana*, ed. by Mary E. Kropp Dakubu, 91–101. London: Kegan Paul International for the International African Institute.
- EFRON, BRADLEY, and ROBERT J. TIBSHIRANI. 1993. *An introduction to the bootstrap*. New York: Chapman and Hall.
- GORDON, RAYMOND G., ed. 2005. *Ethnologue: Languages of the world, fifteenth edition*. Dallas: SIL International. <http://www.ethnologue.com>.
- GREENBERG, JOSEPH H. 1960. A quantitative approach to the morphological typology of languages. *International Journal of American Linguistics* 26:178–194.
- GUDSCHINSKY, SARAH C. 1955. Lexico-statistical skewing from dialect borrowing. *International Journal of American Linguistics* 21:138–149.
- GUDSCHINSKY, SARAH C. 1956. The ABC's of lexicostatistics (glottochronology). *Word* 12:175–210.
- HARRISON, BYRON, ANNETTE HARRISON, and MICHAEL J. RUECK. 1999. Southern Songhay speech varieties in Niger: A sociolinguistic survey of the Zarma, Songhay, Kurtey,

- Wogo, and Dendi peoples of Niger. *SIL Electronic Survey Reports 1999-004*. Dallas: SIL International. <http://www.sil.org/silesr/1999/004/zarmarpt4.pdf>.
- HATFIELD, DEBORAH H., and MICHAEL M. MCHENRY. 1998. A sociolinguistic survey of the Ayizo language area. Ms., SIL International, Cotonou.
- HATFIELD, DEBORAH H., BONNIE J. HENSON, and MICHAEL M. MCHENRY. 1998. A sociolinguistic survey of the Kotafon language area. Ms., SIL International, Cotonou.
- HENSON, BONNIE J., and ANGELA KLUGE. 1999. A sociolinguistic survey of the Xwla language area. Ms., SIL International, Cotonou.
- HENSON, BONNIE J., ERIC C. JOHNSON, and ANGELA KLUGE. 1999. A sociolinguistic survey of the Xwela language area. Ms., SIL International, Cotonou.
- HOCHSTETLER, J. LEE, JUDE A. DURIEUX, EVELIN I. K. DURIEUX-BOON. 2004. Sociolinguistic survey of the Dogon language area. *SIL Electronic Survey Reports 2004-004*. Dallas: SIL International. <http://www.sil.org/silesr/2004/silesr2004-004.pdf>.
- INTERNATIONAL LANGUAGE ASSESSMENT CONFERENCE. 1990. Language assessment criteria: Conference recommendations. In *Proceedings of the Summer Institute of Linguistics International Language Assessment Conference, Horsleys Green, 23–31 May 1989*, ed. by Gloria E. Kindell, 27–29. Dallas: Summer Institute of Linguistics.
- KINDELL, GLORIA E. 1994. A sociology of language approach to assessing potential translation effectiveness in multilingual societies. *Notes on Sociolinguistics* 12:26–29.
- KLUGE, ANGELA. 2000. The Gbe language varieties of West Africa: A quantitative analysis of lexical and grammatical features. MA thesis, University of Wales, College of Cardiff.
- KLUGE, ANGELA. 2005. A synchronic lexical study of Gbe language varieties: The effects of different similarity judgment criteria. *Linguistic Discovery* 3(1):22–53. http://linguistic-discovery.dartmouth.edu/webobjbin/WebObjects/Journals.woa/xmlpage/1/issue?tmpl=ld_issue.
- KLUGE, ANGELA. 2006. Qualitative and quantitative analysis of grammatical features elicited among the Gbe language varieties of West Africa. *Journal of African Languages and Linguistics* 27(1):53–86. <http://www.atypon-link.com/WDG/toc/jall/27/1>.
- KORHONEN, ELSA, MIJRA SAKSA, and RONALD J. SIM. 1986a. A dialect study of Kambaata and Hadiyya (Ethiopia). *Afrikanistische Arbeitspapiere* 5:5–41.
- KORHONEN, ELSA, MIJRA SAKSA, and RONALD J. SIM. 1986b. A dialect study of Kambaata and Hadiyya (Ethiopia). Part 2: Appendices. *Afrikanistische Arbeitspapiere* 6:71–121.
- KRUSKAL, JOSEPH B., and MYRON WISH. 1978. *Multidimensional scaling*. Beverley Hills: Sage Publications.
- MEYERS, JIM, HEATHER MEYERS, BEN RICE, and SUSAN RICE. 2003. Report on the Seruyan river surveys in central Kalimantan: Kohin, Keninjal and Sebaun (Dohoi) languages. *SIL Electronic Survey Reports 2003–017*. Dallas: SIL International. <http://www.sil.org/silesr/2003/silesr2003-017.pdf>.
- NORUŠIS, MARIJA J. 1993. *SPSS for Windows: Professional statistics. Release 6.0*. Chicago: SPSS Inc.
- PROBST, ULRICH. 1992. Multimorphemic words in lexicostatistics: An evaluation on use and limitations of different approaches. Paper presented at the 1992 SIL Survey Course. Ms., Horsleys Green, UK.
- R-Project. n.d. *The R project for statistical computing*. <http://www.r-project.org>.

- RUECK, MIKE, and TIM JORE. 2003. The sociolinguistic situation of the Sam people. *SIL Electronic Survey Reports 2003-019*. Dallas: SIL International. <http://www.sil.org/silesr/2003/silesr2003-019.pdf>.
- SANDERS, ARDEN. G. 1977. Guidelines for conducting a lexicostatistics survey in Papua New Guinea. In *Language planning and survey techniques*. Workpapers in Papua New Guinea Languages 21, ed. by Richard E. Loving and Gary F. Simons, 21–41. Ukarumpa: Summer Institute of Linguistics, 21–41.
- SCHADEBERG, THILO C. 1990. *Lexistat. Version 2.1*. Leiden: Rijksuniversiteit te Leiden, Afrikaanse Taalkunde.
- SIM, RONALD J. 1980. *A sociolinguistic profile of Maasai-Samburu-Ilchamus languages*. Language Data, African Series 16. Dallas: Summer Institute of Linguistics.
- SIMONS, GARY F., and TED G. BERGMAN. 2004. Explain extensibility of literature—Answer keys (Module EE). In *'Principles of language survey' course*, ed. by Gary F. Simons and Ted G. Bergman. Ms., SIL International, Dallas.
- SPSS INC. 1998. *SPSS for Windows: Release 9.0*. SPSS, Inc., 1989–1999.
- STEWART, JOHN M. 1989. Kwa. In *The Niger-Congo languages*, ed. by John Bendor-Samuel, 217–245. Lanham, MD: The University Press of America.
- STEWART, JOHN M. 1994. Review article: The comparative phonology of Gbe and its significance for that of Kwa and Volta-Congo. *Journal of African Languages and Linguistics* 15:175–193.
- SULLIVAN, NANCY, and TERRANCE SULLIVAN. 2000. Information on Aja. [Email] Personal email from Nancy Sullivan and Terrance Sullivan, the SIL team assigned to the Aja language project. (June 11, 2000).
- TĚŠITELOVÁ, MARIE. 1992. *Quantitative linguistics*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- TOMPKINS, BARBARA, and ANGELA KLUGE. 2002. Sociolinguistic survey of the Aja language area. *SIL Electronic Survey Reports 2002-020*. Dallas: SIL International. <http://www.sil.org/silesr/2002/020/SILESR2002-020.pdf>.
- WIESEMANN, URSULA. 1989. Features of grammar in language assessment. Paper presented at the International Language Assessment Conference, Horsleys Green, UK. May 24, 1989.
- WIESEMANN, URSULA, ETIENNE SADEBOUO, and MAURICE TADADJEU. 1983. *Guide pour le développement des systèmes d'écriture des langues africaines*. Collection Propelica 2. Yaoundé: Société Internationale de Linguistique and University of Yaoundé.
- WIESEMANN, URSULA, ed. 1986. *Pronominal systems*. Tübingen: Gunter Narr Verlag.
- WILLIAMSON, KAY, and ROGER M. BLENCH. 2000. Niger-Congo. In *African languages: An introduction*, ed. by Bernd Heine and Derek Nurse, 11–42. Cambridge: Cambridge University Press.
- WIMBISH, JOHN S. 1989. *Wordsurv: A program for analyzing language survey word lists*. Occasional Publications in Academic Computing 13. Dallas: Summer Institute of Linguistics.

Angela Kluge
angela_kluge@sil.org

