Journal of Milev Research and Studies
ISSN : 2392 - 5361
fifth Issue / Juin 2017

Mila Univ center. Publish. Co.

Measuring Chaoui Linguistic Variation in the City of Oum El Bouaghi: A Dialectometric Study

قياس التنوع اللساني في اللهجة الشاوية بمدينة أم البواقي: دراسة ديناليكتوميترية

Bouri Hadj
Belkaied University, Tlemcen Algeria.

Received date: 13/02/2017 *Accepted paper: 04/06/2017*

Abstract :
This study aims at measuring the impact of geographic distance on the linguistic difference. A questionnaire was designed to elicit Chaoui lexis based on the Swadesh list. A levenshtein algorithm was applied to measure the linguistic distance between all the municipalities of the city of Oum El Bouaghi. Also, a correlation experimental design was administered to conduct the study and measure how language varies according to the geographical landscape of the Berber region of the Chaouia. The results show a positive influence of the geographical distance on the Chaoui linguistic diversity and the alternative hypothesis was confirmed. The implementation of inferential statistics along with the levenshtein algorithm helps in understanding how the Berber language intersect with its geographical landscape.
Key words: *dialectometry, dialectology, Berber linguistics, geolinguistics levenshtein distance.*

الملخّص:

تهدف هذه الدراسة إلى قياس أثر المسافة الجغرافية على الاختلاف اللغوي. تم تصميم استبيان لاستخلاص مدونة أمازيغية استنادا إلى قائمة سواديش. تم تطبيق خوارزمية ليفنشتين لقياس المسافة اللغوية بين جميع بلديات مدينة أم البواقي. كما تم استخدام تصميم تجريبي للارتباط لإجراء الدراسة وقياس مدى اختلاف اللغة تبعا للمسافة الجغرافية للمنطقة البربرية في الشاوية. أظهرت النتائج تأثيرا إيجابيا للمسافة الجغرافية على التنوع اللغوي للغة الشاوية، وتم تأكيد الفرضية البديلة. تطبيق الإحصاءات الاستنتاجية جنبا إلى جنب مع خوارزمية ليفنشتين يساعد في فهم كيفية تقاطع اللغة الأمازيغية مع البعد الجغرافي.

كلمات مفتاحية: قياس اللهجي، علم اللهجات، علم اللغة الأمازيغي، مسافة الليفينشتاين، الجغرافية اللسانية.

Introduction

Studies on Berber dialectometry represent a growing field in linguistics and language of the minorities. Both concepts of physical distance and linguistic diversity are central to the study of the impact of geographical features on language variation. Traditionally, Hans Goebl (2008, 2014, 2010, 1982) has subscribed to the belief that language variation according to geographical distance can be measured. Traditionally, linguistics scholars have subscribed to the belief that language boundaries or isoglosses with all its types are geolinguistic zones that can be measured and the amount of linguistic diversity can be delimited. Since the appearance of dialectometry imaginary linguistic boundaries has been subject for further clarification. The impact of geographical distance on language diversity was a key issue in dialectology, traditional linguists tried to study the relationship between language and geography with poor approaches and sketchy methodological steps. Both the choice of the linguistic features as well as the sampling of remote populations subdued inadequacy. However, knowing the importance of the impact of geographical elements in determining linguistic variation is primordial. Results from earlier studies demonstrate a strong and consistent association between geography and language. It has been observed that the larger the geographical distance is the diverse the linguistic features will be.

What we know about language variation and geography comes from accounts by Goebl and many other dialectologists till Peter Trudgill. To date, there has been little agreement about how best to design linguistic atlases and how to approach in scientific research the linguistic diversity in remote geographical areas. Also, there is a current paucity of high-quality research on nonexperimental research in Berber dialectometry. Previous studies have failed to consider the geographical element as an independent variable that modifies considerably the linguistic variation of the Berber language; however, there has been no empirical evidence that clarifies the lexicostatistics of the Chaoui language and its clear distribution on the geographical distance. Previous studies in Chaoui dialectology have suffered from several conceptual and methodological weaknesses. Many sociolinguists from the Maghreb have highlighted linguistic variation broadly and mentioned the isoglosses between different regional varieties either in traditional inaccurate maps, as in the works of André Basset (Chaker, 1995b), or researchers were unable to collect and draw geolinguistic data from the vast and complex mountainous area of the Aures until the vast plains in the South of Constantine. The extent to which geographical distance affects linguistic features of Chaoui is still poorly highlighted by many Dialectologists. In this context, this paper comes to investigate the design and the implementation of lexicostatistical as well as geographical techniques to understand fully how Chaoui of the plains correlates with geographical features of the region. This study seeks to answer the following specific hypothesis: In the city of Oum El Bouaghi, Chaoui ’s lexical features vary due to geographical distance. The null hypothesis is: In the city of Oum El Bouaghi, Chaoui ’s lexical features do not vary due to geographical distance. This study draws on two theoretical frameworks: First a levenshtein algorithm was applied on 105 words list questionnaire based on Swadesh list (Zastrow, 2011) which are “expected to be culturally neutral and stable over time, a real influence is kept to a minimum and diachronic conclusions are potentially justified” (Jack Grieve, 2011). In this questionnaire informants write the equivalent of the word in Arabic in their local Chaoui dialect. Also a non-experimental study was conducted were the researcher sought to find a correlation between geographical distance and linguistic diversity. The experimental work presented here provides one of the first investigations into how to measure

Chaoui language variation according to geographical distance. My personal experience as a researcher in dialectometry and lecturer in sociolinguistics at the University of Larbi Ben Mhidi in the city of Oum El Bouaghi has prompted this research. This study does not engage with all the Chaoui lexis in its details this is why establishing linguistic variation and geographical dimension goes beyond its scope. Throughout this paper, the term dialectometry will refer to "the measurement of dialect differences, i.e. linguistic differences whose distribution is determined primarily by geography" (Nerbonne & Kretzschmar, 2003). In this work we are going to approach the literature related to Berber dialectology in general and the Chaoui linguistic studies in specific with the aim of highlighting the linguistic works and contributions on this language. Then, we are going to advance the main methodological features and approaches to undertake this study. Finally, we are going to discuss the results drawn from this research.

Literature Review:

What we know about Chaoui dialectology is largely based upon traditional empirical studies that investigated and plotted many linguistic and cultural features with geographical maps (Basset, 1883; Chaker, 1995a). The academic literature on Chaoui language is extensive and focuses particularly on this Amazigh variety as a one unified language. The major scholars of Berber studies have never mentioned how Chaoui varies according to either geographical or social dimensions; they have focused on the grammar of the language as an entity rather than as a language with different dialects. Awareness of the Chaoui (Chaouia, Shawia) is not recent, having possibly first been described in report about the Algerian dialects in the school of oriental languages, in 1856 by M, Reinaud. This latter mentioned that the tribes dwelling the mountainous chain in mid Constantine are called Chaouia (Reinaud, 1856). In his book the "Berbers of Algeria", Kimble described how the Chaoui language ". . . owes more to Arab influence than does Kabyle, because they live in closer contact with the nomadic caravan life of the Sahara" (1941). Eventhough Reinaud and Kimble mentioned Chaoui their studies still lack accurate linguistic investigation. The first serious analysis of Chaoui emerged in (1883) in the seminal work of Rene Basset "Notes de Lexicographie Berbere" where he compared the lexis of this variety with the other Amazigh one. Much of the previous research on Chaoui dialectology has been exploratory in nature. Salem Chacker focused

on how the morphological form of noun changes from one variety to another and insisted on its specificity as it ". . . is one of the most delicate points of Berber grammatical system" (Chaker, 1988). Moreover, few recent studies (Lafkioui, 2008a, 2008b) have shown that Amazigh language in Morocco is positively related to geographic distance . Lafkioui's Numerical dialectometry analyses of Rif-Berber lexis where she applied cluster analysis and multi-dimensional scaling found a positive correlation between the independent variable, geographic distance, and the dependent variable, lexical variation. As Lafkioui argues:

> "In terms of linguistic planning, dialectometry can provide a koinè based on the quantitative classification of their linguistic facts according to the criterion of resemblance / difference. This koinè should, however, be backed up by comparative analyzes of a historical and typological type in order to create a true reference representing as closely as possible the linguistic variation of the tariff as well as its uniformity" (2008b).[1]

More recent examples of descriptive statistics in Chaoui language variation can be found in the works of Lounissi Salim (2011) where linguistic charts and tables were presented to display the linguistic variation of Chaoui .

All of the studies reviewed here, either traditional or contemporary, support the hypothesis that North African Berber variation is caused by the vast geographical distance. Besides, the need for an accurate dialectometric study is a necessity to understand fully Amazigh language.

[1] Au plan de l'aménagement linguistique, la dialectométrie peut fournir une koinè moyenne des variétés rifaines fondée sur la classification quantitative de leurs faits lin- guistiques suivant le critère de ressemblance/différence. Cette koinè devrait cependant être épaulée par des analyses comparatives de type historique et typologique afin de créer une véritable référence représentant le mieux plus possible la variation linguistique du tarifit ainsi que son uniformité.

Methods in Dialectometry

Recently, a considerable literature has grown up around the theme of dialectometry in general (Haimerl, 2006; Heeringa, 1970; Mucha & Haimerl, 2005; Nerbonne & Kretzschmar, 2003; Szmrecsanyi, 2008). Dialectometry has been studied extensively since the last decade of the twentieth century and, as a concept in computational linguistics, it is widespread among scholars in northern Europe, Germany the USA and other parts of the world. It is also fundamental to contemporary linguistics since computational tools have given larger perspectives to linguistic studies. Nerbonne and Heeringa are major contributors in this field with their numerous scientific articles. "In dialectometry, the dialect data collected mostly in language lexicon or dialect dictionaries are analysed by means of quantitative methods (statistics, information theory, etc.) with the aid of electronic data processing systems and methods"(Zastrow, 2011). The aim is to make the linguistic structures between the individual dialects of a language visible. The levenshtein algorithm is one of the key components of dialectometry. Evidence suggests that geographical distance is among the most important factors for a diverse language. In recent years, researchers have shown an increased interest in Berber dialectometry. Lafkioui (2008) has been attracting considerable interest since the beginning of 2000. One advantage of using computational approaches to study dialect variation is that it allows the synthetic quantitative analysis and apprehension of linguistic atlas using geolinguistic and numerical taxonomies. Both geolinguistic and statistical calculations are displayed on charts using VDM Visual Dialectometry designed in 2000 by Edgar HAIMERL (Hans Goebl, 2010; Jeszenszky & Weibel, 2015).

Chaoui Dialectometry

Twenty one informants from all the municipalities of the city of Oum El Bouaghi were recruited for this study, as in the map bellow figure 1.

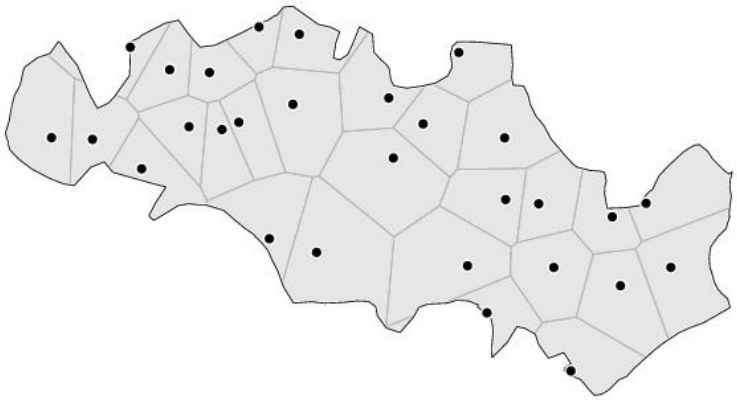


Figure 1: The Main municipalities of the city of Oum El Boughi where the research was conducted

The fundamental criteria for selecting the subjects were as follows:

- Informants are born in the Chaoui region.
- They are fluent speakers of the Chaoui language
- The Chaoui language is spoken at home by all members of the family.
- Both sexes are given the same opportunity to take part in this study.
- Though, very old informants are valued since they preserve the original Chaoui vocabulary and rarely use Arabic loan words, this research avoided NORMS (Non-mobile Old Rural Males) and all the criteria of modern Sociolinguistics were applied.

Data were collected using a questionnaire where the informants were asked to fill the appropriate Chaoui word in front of the equivalent Arab one. The list of 105 words is selected from the Chaoui vocabulary under the criteria set by Swadesh list, as aformentioned. The method applied in this study in String Edit Distance Tokenized. The local incoherence is 0.36 which means that the results drawn from these municipalities are valuable and reliable because the Lower the values for Local incoherent are the better the results will be. Also the Cronbach's Alpha of the questionnaire is 0.39 percent which means that the validity of the questionnaire is high and can be trusted as a tool of collection and measurement of the linguistic data aggregated.

1.1 Levenshtein distance

The collection of data was conducted over the course of the growing period of the second semester of 2016. All the work on the computer was carried out using R (Team, 2017) software for statistical analysis and GABMAP (John Nerbonne, Rinke Colen, Charlotte Gooskens, Peter Kleiweg, 2011) an online software for dialectal data mining and visualization. In order to understand how geographical distance regulates linguistic variation, a levenshtein algorithm was applied to compare between the linguistic strings: where addition, omission or substitution of sounds were given the value of 1 between each location and the rest of the other municipalities at the level of each string that is to say each lexical item in used as a basis to compare between all the municipalities of the city of Oum El Bouaghi. As Tables 1 and 2 display an example of how the operation processes at the both levels: lexical and syntaxic form of the Berber word *camel* and the sentence: *I am older than you.* This comparison is made between four Municipalities where the linguistic distance between Ain Babouche and Ouled Guecem is 3 and Ain Babouche and Berriche at the syntactic level is 5. This function was repeated with all the 105 lexemes in a binary comparison between all the 29 municipalities.

Table 1. Binary distance matrix of the lexeme "Camel" between Ain Bebouche and Ouled Guecem municipalities

Ain Babouche — Ouled Guecem

a	l	a	g	h		m	i	
a	l		g	h	a	m		
		1			1		1	3

Table 2. Syntactic distance of the sentence **I am older than you** between Ain Babouche and Berriche municipalities

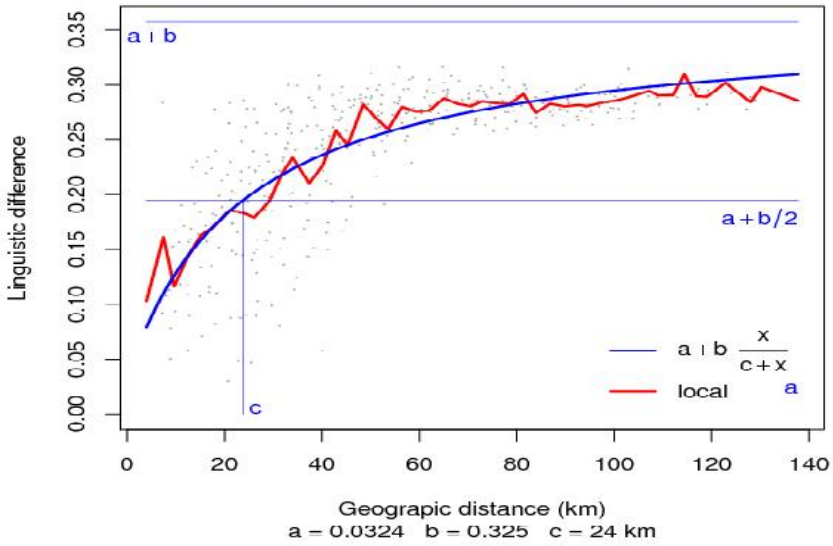
Ain Babouche — Berriche

n	a	t	c	h	SP	A	k		t	h	a	r	SP	m	a	n	n	a	k	
n	a	t	c	h	SP		k	a	t	h		r			a	n	n	a	k	
						1		1			1		1	1						5

Correlation

Another statistical analysis was used based on nonexperimental studies a correlation experimental design was administered. The aim from this latter is to interpret how linguistic data are influenced by geographical distance. A plot with local regression and asymptotic regression was designed and the value seen on the plot chart shows the value of a: 0.03243 and b: 0.32503. The value of these two numbers is very small which indicates that there is a very low signal ratio in the data. Also the value of c equals 23.87622 which mean that linguistic variation is measurable over a large geographic distance. These statistical results can be clearly seen on figure 2 where linguistic difference is plotted with geographic distance. After that, a linguistic distance matrix was designed were the language distance is clearly seen, as table 3 shows.

Figure 2. Plot of Geographic distance with linguistic difference



Dendrogram

An MDS analysis is to provide a visual representation of the linguistic distances among the municipalities. The pair-wise aggregate linguistic distances analysis between all sites are analysed to give birth to a chart as seen on figure 2. As seen on the figure it is apparent that the city of Oum El

Bouaghi comprises three main linguistic groups each has its own syntactic and lexical uniformity even though mutual intelligibility remains between both the geographic and ethnic communities.

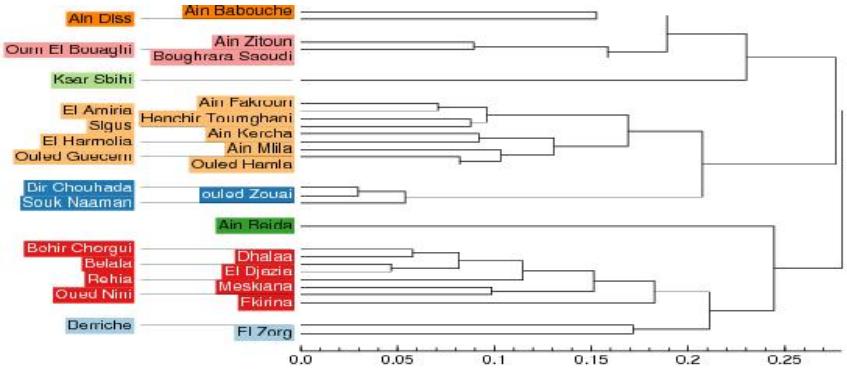


Figure 3: Dendrogram showing the three linguistic groups in the city Oum El Bouaghi

The first question in this study sought to determine how geographic distance determines linguistic difference. The results of this study indicate that to an adequate extent language variation is determined by the length of distances between the different municipalities of the city of Oum El Bouaghi. There are, however, other possible explanations for the regression seen on the correlation of the results is that they may be due to other intervening variables as the ethnic elements or due to historic variable that we failed to control. This is why these findings may be somewhat limited by cultural and historical factors. In the future research, research questions that could be asked should include the historical and ethnic variables that are key elements in the general sociocultural factors of the region.

Conclusion

Returning to the question posed at the beginning of this study, it is now possible to state that geographic distance has a great impact on the linguistic difference. Although this study focuses on the geographical element as an independent variable, the findings may well have a bearing on sociocultural aspects or historical ones. The empirical findings in this study provide a new understanding of how Chaoui language diversification is assisted by the vast geographical landscape of the plain region of the Chaouia. Notwithstanding the relatively limited sample and the very short lexeme list, this work offers

valuable insights into the importance the geolinguistic studies of Berber in general and Chaoui language in specific. The major limitation of this study is the focus on geographical variable and the total neglect of sociocultural issues of the region. A key strength of the present study was the integration of the levenshtein algorithm along with implementation of correlation nonexperimental design. It is recommended that further research be undertaken in the historical and social areas.

List of Reference:

Basset, R. (1883). *Notes de Lexicographie Berbere*. (E. Lerous, Ed.), *Journal Asiatique*. Paris.

Case, A. R., & Dialectometry, O. F. (2009). New insights into the use of vdm: some preliminary stages, *2*, 23–35.

Chaker, S. (1988). L â€™ Etat D â€™ Annexion Du Nom. In *Encyclopédie berbère* (Vol. 4, pp. 686–695).

Chaker, S. (1995a). Derivation (linguistique). In *Encyclopédie berbère* (Vol. XV, pp. 1–2).

Chaker, S. (1995b). Dialecte. In *Encyclopédie berbère* (Vol. XV, pp. 1–5).

Goebl, H. (2008). Brève Introduction Aux Problèmes Et Méthodes De La Dialectométrie. *Revue Roumaine de Linguistique*, *1–2*, 87–106.

Goebl, H. (2014). L ’ Impact De La Polynymie Des Cartes D ’ Atlas Sur Le Résultat. *Linguistique Romane et Linguistique Indoeuropéenne*, 243--260.

Goebl, H. (1982). Atlas, Matrices Et Similarities: Petit Apercu

Dialectometrique. *Computers and the Humanities, 16*, 69–84.

Goebl, H. (2010). Dialectometry: Theoretical Prerequisites, Practical Problems, And Concrete Applications (Mainly With Examples Drawn From The "Atlas Linguistique De La France", 1902-1910). *Dialectologia. Special Issue, 1*, 63–77.

Haimerl, E. (2006). Database Design and Technical Solutions for the Management , Calculation , and Visualization of Dialect Mass Data. *Literary and Linguistic Computing*, *21*(4), 437–444. https://doi.org/10.1093/llc/fql037

Heeringa, W. (1970). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. University of Groningen.

Jack Grieve. (2011). The use of spatial autocorrelation statistics for the analysis of regional linguistic variation. In A. Z. and A. Lüdeling (Ed.), *Proceedings of Quantitative Investigations in Theoretical Linguistics 4* (pp. 34–36). Humboldt-Universität zu Berlin.

Jeszenszky, P., & Weibel, R. (2015). Measuring boundaries in the dialect continuum. *Proceedings of the AGILE*.

John Nerbonne, Rinke Colen, Charlotte Gooskens, Peter Kleiweg, and T. L. (2011). Gabmap — A Web Application for Dialectology. *Dialectologia Special Issue II*.

Kimble, G. H. T. (1941). The Berbers of Eastern Algeria. *The Geographical Journal*, *97*(6), 337–347. Retrieved from http://www.jstor.org/stable/1788169 Accessed:

Lafkioui, M. (2008a). Dialectometry Analyses Of Berber Lexis. *Folia Orientalia*, *44*, 71–88.

Lafkioui, M. (2008b). Pour la démarche géolinguistique de la standardisation des variétés amazighes du Rif. *Afrika Focus*, *21*(1), 97--102.

Mucha, H., & Haimerl, E. (2005). Automatic Validation of Hierarchical Cluster Analysis with Application in Dialectometry. In *Classification*

the Ubiquitous Challenge (pp. 513--520). Springer.

Nerbonne, J., & Kretzschmar, W. (2003). Introducing Computational Techniques in Dialectometry. In *Computers and the Humanities* (Vol. 37, pp. 245–255). Netherlands.: Kluwer Academic Publishers.

Reinaud, M. (1856). *Rapport sur le Tableau des dialectes de l'Algérie et des contrées voisines*. Paris.

Salim, L. (2011). *Etude de Geogrphie Linguistique Chaoui sur les Plans Phonetico-phonologique et Lexical*. Mouloud Mammeri De Tizi Ouzou.

Szmrecsanyi, B. (2008). Analyzing aggregated linguistic data, 1–27.

Team, R. C. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Zastrow, T. (2011). *Neue Analyse- und Visualisierungsmethoden in der Dialektometrie*. akultät der Eberhard Karls Universität.