

Dialectology and linguistic geography

Mena Lafkioui

Ghent University – Università di Milano-Bicocca

I. Introduction

Both dialectology and linguistic geography study language variation, change and diffusion. The former examines geolinguistic data synchronically and diachronically on an intragenetic level (e.g. Bantu, Berber) or an intergenetic phylum level (e.g. Niger-Congo, Afroasiatic), whereas the latter focuses on stable, contact-induced linguistic changes on an intergenetic interphyla level in adjacent areas.¹

In the present chapter, I address these linguistic disciplines from an African perspective, explaining their fundamental principles by means of specific African-related cases. Furthermore, I compare the different approaches to African areal linguistics, pointing out their respective potentials and limitations.

This introduction is followed by a preliminary explanation of the dialectological framework, problematizing some of its theoretical and methodological concepts, such as ‘dialect consciousness’, ‘interintelligibility’, ‘language continuum’ and the ‘language versus dialect’ dichotomy (section II). In the same section, a detailed dialectometry classification and its various methods are presented. In section III, a critical overview of the existing studies in African areal classification is given. The chapter ends with a number of conclusions about language variation and its connection with language diffusion and evolution with respect to the African area (section IV).

II. Dialectology

Dialectology considers variation in language based most commonly on geographical distribution and its correlated phonological, grammatical and lexical features. Synchronic dialectology studies the synchronic linguistic proximity of geographically contiguous language varieties. Diachronic, or historical, dialectology describes and explains by means of sociohistorical reconstructions the language evolution and diffusion processes which led to the current variation in form, meaning and structure. The two approaches are interrelated and therefore no clear-cut distinction can be made between them. Indeed, in variational linguistic research (viz. dialectology), any geographical diversity necessarily refers to a temporal diversity (de Saussure 1973 [1916]; Lafkioui 2007). In addition to the horizontal (geographical) and vertical (historical) dimensions, linguistic variation also involves social diversity and individual diversity. The former includes parameters such as habitat (rural/urban), country of origin/country of immigration, gender, ethnicity, age, social class, educational level and educational type; the latter overlaps with social variation but also incorporates ideological and psychohistorical aspects of the

¹ The term genetic is used here in the sense of genealogical inheritance.

individual speakers (Lafkioui 2007, 2011a). In the case of Africa, it is historical dialectology which, of the two sub-disciplines, is the one in greater need of additional and advanced study. A notable exception to this lack is the research conducted in creolistics, which benefits not only dialectological but also linguistic geographical studies (e.g. Gilman 1986; Salikoko 2000).²

Although the terms ‘dialectology’ and ‘dialect’ do not involve any value judgements in the eyes of dialectologists, they may be subject to controversy, in particular when the latter is used in contrast with the concept of a ‘language’ in sociolinguistic situations where linguistic varieties (lects or geolects) are in sociopolitical conflict, especially as the language-dialect distinction cannot be established on scientific criteria.

Concepts like ‘dialect consciousness’ and ‘interintelligibility’, even though often used as defining parameters (e.g. Daan 1999), do not suffice to distinguish languages from dialects or dialects from each other. Neither do they form adequate geographical demarcation devices for the varieties in question (Lafkioui 2007, 2011a), mainly because they are based on the subjective experience of speakers whose values – which are difficult to quantify – depend heavily on their personal disposition (ideological, social and historical). As regards dialect consciousness, even if the speakers regard themselves belonging to the same linguistic variety or group of varieties – calling themselves by the same anthroponym for example – their language attitudes often seem to be motivated by the social status accorded to the variety in question and the degree of in-group solidarity (2007). Moreover, the validity of the interintelligibility criterion is primarily determined by the conscious or unconscious willingness of speakers, which does not always permit the researcher to decide whether mutual comprehension actually stems from no more than their linguistic competencies. It is not uncommon to find speakers who claim not to understand because they refuse to understand (Chambers & Trudgill 1980: 4). Therefore, these two extralinguistic methodological tools should be employed with caution when delimitating linguistic varieties and selecting (field) survey points for this purpose; their use should be confined to subsidiary verifications of the data and analyses drawn from linguistic inventories.

The prime objective of research in dialectology is to analyse and classify the language variation of a specific geographical zone (diatopic variation) and so to identify the causal linguistic boundaries (isoglosses) of the different encountered varieties and variety groups, even when these cut across administrative and thus political borders. Essential to this approach is the construction of a linguistic atlas of the area being studied; that is, a structured set of geolinguistic maps displaying the precise distribution in space of particular linguistic phenomena. As discussed in Lafkioui (2007), a linguistic atlas – especially when it deals with non or little documented and studied varieties – should comprise, and give equal consideration to, all linguistic domains, namely phonetics, phonology, morphology, syntax and lexicon. Such an approach facilitates further linguistic studies such as dialectometry classification, a quantitative methodology for measuring dialectal proximity or distance which is one of the most appropriate developments of this research perspective and is explained in more detail in the next section.

A major problem for classification (both quantitative and qualitative types), and particularly for statistical classifications such as dialectometrics, is the existence of so-called ‘dialect

² For African historical dialectology, see e.g. Vossen (1988, 2008), Sommer & Vossen (1993) and Lafkioui (2007, 2009c, 2011a, 2011b).

continua'. These impede a clear-cut demarcation of linguistic varieties (dialects and languages) because the progressive change which they exhibit prevents convergence to precise linguistic border lines (isoglosses and bundles of isoglosses). Well-known examples of African dialect continua are the Bantu continua (e.g. Guthrie 1967-71; Möhlig 1979; Heine & Möhlig 1980), one of which is that found in Kenya, where the status of 'language' has been assigned to varieties such as Gikuyu, Meru and Kamba mainly on the basis of non-linguistic parameters (Heine & Möhlig 1980: 9-10).³ Another case of dialect continua is the Berber one (Basset 1929, 1959; Lafkioui 2007), of which the Rif-Berber continuum (North Morocco) has been the subject of a thorough dialectometry classification that is demonstrated in the following section (see also Lafkioui 2008, 2009a, 2009b).

1. Introduction to methods in dialectometry

Dialectometry is a quantitative methodology for calculating linguistic distances between linguistic varieties. The most frequently used dialectometry methods can be divided into the categories of traditional and computational methods.⁴

The most well-known traditional approaches are those based on the concept of isogloss, which is a line that bisects a geographic map into separate zones according to the detected linguistic features. The classification of the varieties is deduced from the arrangement of isoglosses, clusters of isoglosses (Goossens 1969) or clusters of demarcative isoglosses (Stankiewicz 1957; Garde 1961) on the geolinguistic map.⁵ Although this method allows for verification of the visualised facts, it has several disadvantages, including the difficulty to find clusters of isoglosses that precisely divide the geolinguistic area examined (Kessler 1995).⁶

Another traditional technique is the geolinguistic structuring method which divides a geographic area depending on the linguistic structure of its varieties (Moulton 1960; Goossens 1965, among others). For instance, varieties with the same phonemic system are part of the same

³ Nevertheless, Möhlig (1974, 1980) and Heine (1971, 1974) were the first to apply statistical methods to African corpora; the former in order to get a dialectometry classification of Bantu varieties, whereas the latter is interested in lexicostatistics and its importance to African historical classification. See also Bastin (1983) for a lexicostatistic study of Bantu.

⁴ There also exist different *perceptual approaches* that permit to draw sociolinguistic borders based on the speaker's dialectal conscience (Weijnen 1946, 1966; Rensink 1955; Daan & Blok 1969; Goossens 1997; among others).

⁵ The qualifying term 'demarcative', added to the common dialectology criterion of 'isogloss clusters' (Goossens: 1969, 54), refers to the structural value of the isoglosses relating to the material aspect of the phenomena as well as to their relative distribution (direction and density). Thus, not only the quantitative dimension (number) of isoglosses is relevant to the typology of classification, but also the qualitative aspect, i.e. their degree of importance. However, non-demarcative isoglosses may also be of great significance for the classification, especially when they allow for an evaluation of the results. On the relationship between structuralism and dialectology, see Forquet (1956), Weinreich (1954), Grosse (1960) and Martinet (1972), among others.

⁶ A significant critique on this method is that it cannot completely exclude some subjectivity because isoglosses might be chosen, a priori, according to the linguistic borders they yield (Goossens 1977).

Mena Lafkioui, Dialectology and linguistic geography, in: Vossen (ed.), Handbook of African Languages, Oxford: Oxford University Press, forthcoming in 2012.

geolinguistic group. However, classifications based on this method are mainly phonological and therefore lack an interpretation basis that is connected with other linguistic dimensions.⁷

The computational dialectometry methods are numerous and are currently considered most adequate for reasons I will explain later on in this section. The foundations of digital automated dialectometry were established by Séguy (1973) with his analytical method to calculate the linguistic differences between varieties of Gascogne. The comparison is based on an algorithm which classifies data as identical or non-identical. The sum of the measured distances between two varieties matches their linguistic distance. The visualisation of the classification analysis is conducted through lines of various types (bold/non bold, dotted/non dotted, etc.), which divide the region according to the linguistic differences of the varieties. As a counterpart, Goeble (1982, 1993) has calculated the similarities between varieties from Italy and Southern Switzerland. Even though the results of the calculation of Séguy and Goeble have the merit of being objective, they lack refinement because their technique excludes distance graduation.

The main computational methods based on the frequency of linguistic variants are the “Corpus Frequency Method” (Hoppenbrouwers & Hoppenbrouwers 1988, 2001) and the “Frequency per Word Method” (Nerbonne & Heeringa 1998, 2001). The basic principle of the first approach is that the degree of difference/similarity between two varieties is derived from comparing the frequency of the marked linguistic features of their variants. The problem in this approach is that the entity ‘word’ is not considered as a linguistic unit. However, this obstacle is removed by the second approach which assigns to words the status of ‘units’ functioning as such. Nevertheless, the two classification tools do not take into account the order of the phonic units in the sequence.

The “Levenshtein distance” (Lv), on the other hand, allows incorporating the parameter of sequential ordering of phonic units in the classification, which makes it more appropriate than other digital/numerical methods. This tool has been introduced in dialectometry by Kessler (1995), who has applied it to a corpus of Irish Gaelic. The Levenshtein distance measure corresponds to the numerical value of the lowest cost of operations (insertions, deletions and substitutions) needed to convert a string of characters into another (Kruskal 1999). One of the most employed techniques of comparison is the “phone string comparison” in which all operations have the same cost, regardless of the degree of affinity between the phonic units: the pair [t, d] has the same cost as the pair [u, t] and [u, u:]. Yet, with the technique of “feature string comparison” phonetic features of phonic units can be compared: the cost of the pairs [u, t] and [u, u:] is not equal because the phonetic affinity between the phonic units of [u, u:] is greater than that of [u, t].

⁷ Although the frequency of the compared variants is taken into account (Kocks 1970), this approach does not seem to be the most appropriate (Heeringa 2004: 24-25).

2. Dialectometry analysis and African languages: the case of Berber lexis

Among the different existing dialectometry approaches, I prefer the computational methods because they allow handling large data corpora with certain ease, while ensuring the accuracy and consistency of the analyses. These aims can be achieved thanks to the fact that:

- Distances and frequencies are measured automatically.
- Data are classified digitally.
- Mapping can be assisted by the computer.
- Statistical analyses can be made and displayed automatically.

The dialectometry analyses that I present in this section were performed with the free software of Kleiweg (RuG/L04).⁸ In order to complete a displayed dialectometry analysis, all the procedural steps summarised below are indispensable:

Table 1: General procedure of computational dialectometry analysis

| | |
|---------------|----------------------------------------------|
| Step 1 | Linguistic Atlas = georeferenced data source |
| Step 2 | Data Matrix |
| Step 3 | Distance Matrix |
| Step 4 | Analysis |
| Step 5 | Visualisation |

2.1. The Linguistic Atlas of the Rif as a data source

The Rif is that region of North Morocco stretching from the Strait of Gibraltar in the West to the Algerian frontier in the East. The Rif-Berber varieties (Tarifit) belong to the northern Berber languages and thus are part of the large Afroasiatic language phylum. The Berber-speaking area of the Rif is delimited:

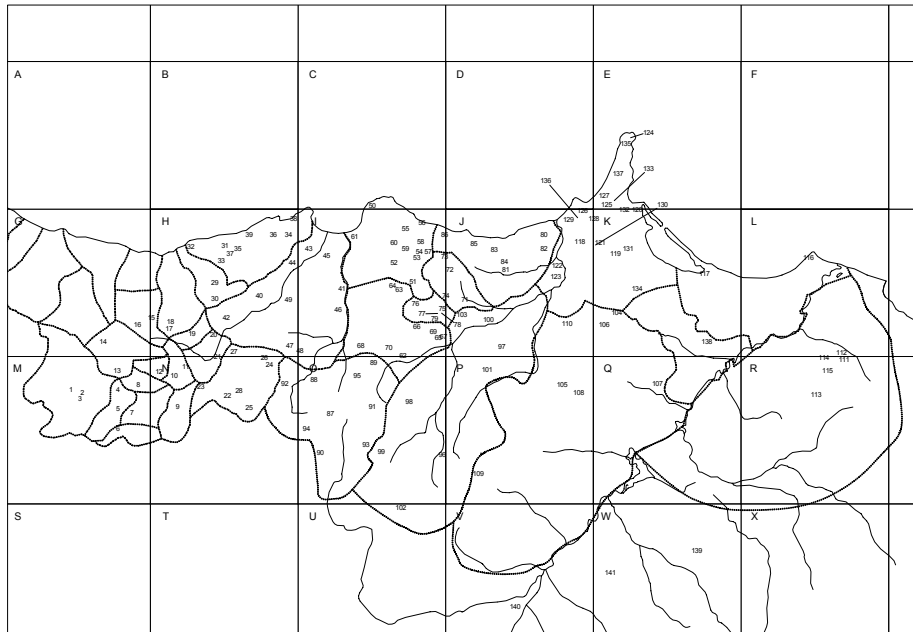
- In the West, by the varieties of the Ktama tribe, (the so-called Senhaja varieties).
- In the South, by the koinè of Gersif, which is the ultimate geographic point where Rif-Berber (Tarifit) is spoken before reaching the corridor of Taza.
- In the East, by the varieties of Iznasen, which have spread to the regions of Arabic speaking varieties to the Morocco-Algerian border.

The lexical data which are compared and classified in this study are collected from the *Atlas linguistique des variétés berbères du Rif* (Lafkioui 2007) or ALR. The digital data corpus consists of sixty-two lexemes regarding the human body (maps 295 to 315), kinship (maps 316 to 321 cards), animals (maps 322 to 327), colours (maps 328 and 329), numbers (maps 330 to 332), besides a subset of various nouns and verbs (maps 333 to 356). Of these lexemes, eleven have only one variant per variety; all fifty-one other lexemes display the co-occurrence of multiple variants for each lexeme.

⁸ [Http://odur.let.rug.nl/~kleiweg/L04](http://odur.let.rug.nl/~kleiweg/L04).

Due to the completion of the automated ALR, the data obtained from it are already in digital format, which has avoided a great task of digitising. However, an adaptive conversion to the software RuG/L04 (Kleiweg) was necessary. The ALR also offers a precise digital map of the Rif region (see Figure 1), which is essential to the visualisation of the dialectometry analyses, except for the dendrogram.

Figure 1: Map of the georeferenced survey points of the Rif (Lafkioui 2007)



One hundred forty-one georeferenced points – belonging to thirty-two Rif tribes – were selected from a group of four hundred fifty-two localities in the Rif according to their degree of linguistic variation (Lafkioui 2007).⁹

2.2. Data matrix of Rif-Berber lexis

The data matrix is composed of digital lexical excerpts from the ALR (Lafkioui 2007) converted following the format of the software RuG/L04 (Kleiweg). Here below, a small sample in digital format of the ALR (Mapinfo Professional format; Table 2):

⁹ The survey points were selected on the basis of the principle of equidistance dividing the inquiry field into several grids to which were assigned points that could match with localities on the field. The greater the variation was, the more the grids were reduced. The four hundred fifty-two locations selected for this research were for the most part chosen so that they could, a priori, indicate linguistic borders. This selection mainly stemmed from the scientific and empirical knowledge of the investigator on the different varieties spoken in the Rif area.

Mena Lafkioui, *Dialectology and linguistic geography*, in: Vossen (ed.), *Handbook of African Languages*, Oxford: Oxford University Press, forthcoming in 2012.

Table 2: Data excerpt in digital format of ALR

| SECTO | TRIBE | FULL_NAME_NO | LF461 | LF462 | LF463 | LF464 | LF465 | LF466 | LF467 | LF468 | LF469 | LF470 |
|-------|-------------|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | Ktama | Asammer | 31 | 51 | 52 | 32 | 54 | 51 | 32 | 33 | 53 | 52 |
| 1 | Ktama | Lmexzen | 31 | 51 | 52 | 32 | 54 | 51 | 32 | 33 | 53 | 52 |
| 1 | Ktama | Ssahel | 31 | 51 | 52 | 32 | 54 | 51 | 32 | 33 | 53 | 52 |
| 2 | Taghzut | Lqel'a | 31 | 54 | 52 | 32 | 54 | 51 | 34 | 33 | 54 | 52 |
| 2 | Taghzut | Ssaqya | 31 | 54 | 52 | 32 | 54 | 51 | 34 | 33 | 54 | 52 |
| 3 | Ayt Bucibet | Tarya | 31 | 54 | 52 | 32 | 54 | 51 | 32 | 33 | 54 | 52 |
| 4 | Ayt Hmed | Mazuz | 31 | 51 | 52 | 32 | 54 | 51 | 32 | 33 | 54 | 52 |
| 5 | Ayt Bunsar | Luta | 31 | 32 | 52 | 32 | 54 | 51 | 32 | 33 | 51 | 52 |
| 6 | Ayt Bcir | Tizirt | 31 | 51 | 52 | 32 | 52 | 51 | 33 | 31 | 51 | 52 |
| 7 | Zerqet | Aghennuy | 31 | 32 | 52 | 32 | 52 | 51 | 33 | 31 | 51 | 52 |
| 7 | Zerqet | Wversan | 31 | 32 | 52 | 32 | 52 | 51 | 33 | 31 | 51 | 52 |
| 8 | Ayt Xennus | A'raben | 31 | 32 | 52 | 32 | 54 | 51 | 32 | 33 | 53 | 52 |
| 9 | Ayt Seddat | Azila | 31 | 32 | 52 | 32 | 54 | 51 | 32 | 33 | 53 | 52 |
| 9 | Ayt Seddat | Tamadda | 31 | 32 | 52 | 32 | 54 | 51 | 32 | 33 | 53 | 52 |
| A | Ayt Gmil | Azru n tili | 53 | 53 | 53 | 32 | 51 | 54 | 31 | 52 | 53 | 53 |
| A | Ayt Gmil | Tizi | 53 | 53 | 53 | 32 | 51 | 54 | 31 | 52 | 53 | 53 |
| B | Ayt Bufrah | Igzennayen | 53 | 32 | 52 | 32 | 51 | 13 | 31 | 52 | 53 | 52 |
| B | Ayt Bufrah | Iharunen | 53 | 32 | 52 | 32 | 51 | 13 | 31 | 52 | 53 | 52 |
| C | Targist | Ayt 'Azza | 54 | 52 | 53 | 53 | 51 | 12 | 53 | 52 | 53 | 53 |
| D | Ayt Mezduy | Bni Budjay | 53 | 53 | 53 | 32 | 51 | 54 | 31 | 52 | 53 | 53 |
| D | Ayt Mezduy | Bu'di | 53 | 32 | 52 | 32 | 51 | 51 | 31 | 52 | 53 | 52 |

2.3. Distance matrix of Rif-Berber lexis

This section contrasts the three most employed digital comparison techniques: the Binary distance (Hamming algorithm), the Gewichteter Identitätswert distance (Weighted identity value), and the Levenshtein distance. I will apply these techniques on the Rif-Berber lexical corpus to test their validity and to select the most appropriate to Berber. Each distance measuring allows acquiring precise numerical values derived from the linguistic comparison between the varieties of the Rif area. These values make up the distance matrices (symmetric matrices $N \times N$, N = sum of varieties), whose configuration differs depending on the adopted algorithm.

2.3.1. Binary distance

The Binary distance (Bin) is used to classify lexical units as being identical or non-identical: comparison of type 0-1; 0= resemblance and 1= difference. Table 3 presents an excerpt from the Binary distance matrix of the lexeme 'heel' (ALR, map 312):

Table 3: Excerpt from the Binary distance matrix of the lexeme ‘heel’

| | Tizirt | Aghennuy | Wersan | A'rabén | Azila | Tamadda | Azru n tili | Tizi |
|-------------|--------|----------|--------|---------|-------|---------|-------------|------|
| Wersan | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| A'rabén | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Azila | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Tamadda | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Azru n tili | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| Tizi | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| Igzennayen | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| Iharunen | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| Ayt 'Azza | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

2.3.2. Gewichteter Identitätswert distance

The Gewichteter Identitätswert distance (GIW) deviates from the Binary distance in that the frequency of the lexical variants is considered in the comparison: low-frequency variants weigh heavier than high-frequency variants. The distance obtained by this technique varies between 0 and 1; $\{0 \leq d \leq 1\}$. Table 4 presents an extract from the distance matrix of the lexeme ‘heel’:

Table 4: Excerpt from the GIW distance matrix of the lexeme ‘heel’

| | Tizirt | Aghennuy | Wersan | A'rabén | Azila | Tamadda | Azru n tili | Tizi |
|-------------|-----------|-----------|-----------|-----------|-----------|-----------|-------------|----------|
| Wersan | 0.0501792 | 0.0501792 | 0 | 0.0501792 | 0.0501792 | 0.0501792 | 1 | 1 |
| A'rabén | 0.0501792 | 0.0501792 | 0.0501792 | 0 | 0.0501792 | 0.0501792 | 1 | 1 |
| Azila | 0.0501792 | 0.0501792 | 0.0501792 | 0.0501792 | 0 | 0.0501792 | 1 | 1 |
| Tamadda | 0.0501792 | 0.0501792 | 0.0501792 | 0.0501792 | 0.0501792 | 0 | 1 | 1 |
| Azru n tili | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0.215054 |
| Tizi | 1 | 1 | 1 | 1 | 1 | 1 | 0.215054 | 0 |
| Igzennayen | 1 | 1 | 1 | 1 | 1 | 1 | 0.215054 | 0.215054 |
| Iharunen | 1 | 1 | 1 | 1 | 1 | 1 | 0.215054 | 0.215054 |
| Ayt 'Azza | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

2.3.3. Levenshtein distance

The distance values resulting from a Levenshtein-based comparison – an algorithm taking into account the sequential order of phonic units composing lexemes – fluctuate between 0 and 1, $\{0 \leq d \leq 1\}$, as shown in the following excerpt:

Table 5: Excerpt from the Lv distance matrix of the lexeme ‘heel’

| | Tizirt | Aghennuy | Wersan | A'raben | Azila | Tamadda | Azru n tili | Tizi |
|-------------|----------|----------|----------|----------|----------|----------|-------------|----------|
| Wersan | 0 | 0 | 0 | 0 | 0 | 0 | 0.6 | 0.6 |
| A'raben | 0 | 0 | 0 | 0 | 0 | 0 | 0.6 | 0.6 |
| Azila | 0 | 0 | 0 | 0 | 0 | 0 | 0.6 | 0.6 |
| Tamadda | 0 | 0 | 0 | 0 | 0 | 0 | 0.6 | 0.6 |
| Azru n tili | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0 | 0 |
| Tizi | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0 | 0 |
| Igzennayen | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0 | 0 |
| Iharunen | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0 | 0 |
| Ayt 'Azza | 0.555556 | 0.555556 | 0.555556 | 0.555556 | 0.555556 | 0.555556 | 0.111111 | 0.111111 |

These values result from the selection of the least costly calculation to transform a lexical unit – as a string of phonic units – into another. Table 6 depicts the lowest cost of operations which allow modifying the string *awrez* (heel) into *inerz* (heel):

Table 6: Cost of operations allowing modification of *awrez* into *inerz* (heel)

| | | a | w | r | e | z |
|---|-----|-----|-----|-----|-----|-----|
| | 0 | 0.5 | 1 | 1.5 | 2 | 2.5 |
| i | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 |
| n | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 |
| e | 1.5 | 2 | 2.5 | 3 | 2.5 | 3 |
| r | 2 | 2.5 | 3 | 2.5 | 3 | 3.5 |
| z | 2.5 | 3 | 3.5 | 3 | 3.5 | 3 |

The lowest cost of operations amending *awrez* into *inerz* is 3, which implies that the distance between these two lexemes is 3/5 (5 being the total number of features); consequently, the Levenshtein distance is 60%. These calculations are based on operations that cost 0.5 for an insertion or deletion and 1 for a substitution. Table 7 illustrates this calculation technique:

Table 7 : Example of calculation of Lv distance for modifying *awrez* into *inerz* (heel)

| Tamadda | a | w | r | e | | z | |
|-------------|---|---|-----|---|-----|---|-------------------------|
| Tizi | i | n | | e | r | z | |
| Lv Distance | 1 | 1 | 0,5 | 0 | 0,5 | 0 | 3/5 * 100 = 60 % |

2.4. Numerical dialectometry analyses of Rif-Berber lexis

From the distance matrices, numerical comparative analyses of Berber lexis can be accomplished through two techniques: *Cluster Analysis* and *Multidimensional scaling*. The technique of Cluster Analysis (CA) consists of regrouping data by reducing the distance matrix by means of various algorithms. According to Kleiweg (RuG/L04), I have implemented the Ward algorithm (minimum variance), which is generally regarded as one of the most appropriate algorithms for this type of analysis. On the other hand, multidimensional scaling (MDS) is:

Mena Lafkioui, *Dialectology and linguistic geography*, in: Vossen (ed.), *Handbook of African Languages*, Oxford: Oxford University Press, forthcoming in 2012.

“[...] a technique that, using a table of differences, tries to position a set of elements into some space, such that the relative distances in that space between all elements corresponds as close as possible to those in the table of differences.” (Kleiweg, RuG/L04).

2.5. Visualisation of dialectometry analyses of Rif-Berber lexis

Classification by clustering (CA) necessarily uses a dendrogram for its display. A dendrogram is a complex ranking structure, usually in colour, whose branches represent the linguistic varieties. It can be matched with a digital map, resulting in a geolinguistic map that shows the distribution of linguistic varieties depending on the linguistic differences and the selected classification criteria. In contrast, analyses by Multidimensional Scaling (MDS) directly offer maps on which the relative linguistic variation is gradually represented by different colours.

Here below the lexis classification obtained through L_v distance corresponding with an asymmetrical configuration of 7 clusters which are structured into 2 major clusters distanced from one another by 8.08 (Figure 2). The matching dendrogram shares the same linguistic main delimitation (between groups 6 and 3-4) with the other dendrograms based on the Bin and GIW algorithm (for their visualisation, see Lafkioui 2008, 2009b). This observation is corroborated by the CA_{L_v} maps displayed in Figure 3, of which the 2-cluster map clearly indicates the most distinctive linguistic boundary. It is important to note that the CA_{L_v} map (Figure 2) shows a distribution of the varieties similar to the CA_{Bin} distribution (Lafkioui 2008, 2009b), even though the composition of their respective dendrogram is divergent.

Figure 2: Dendrogram vs. CA Map - L_v – All lexis

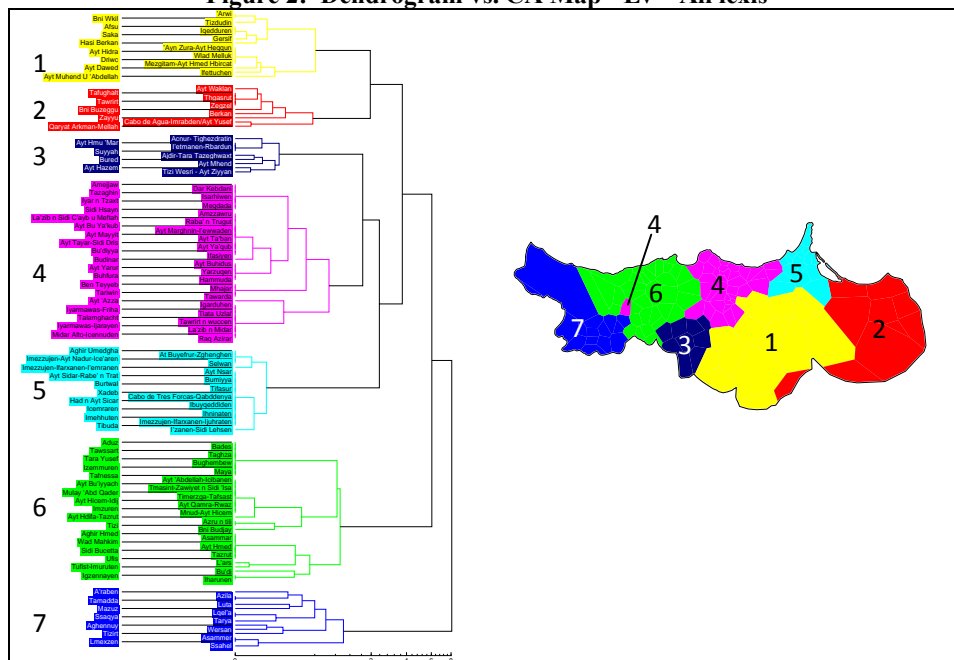
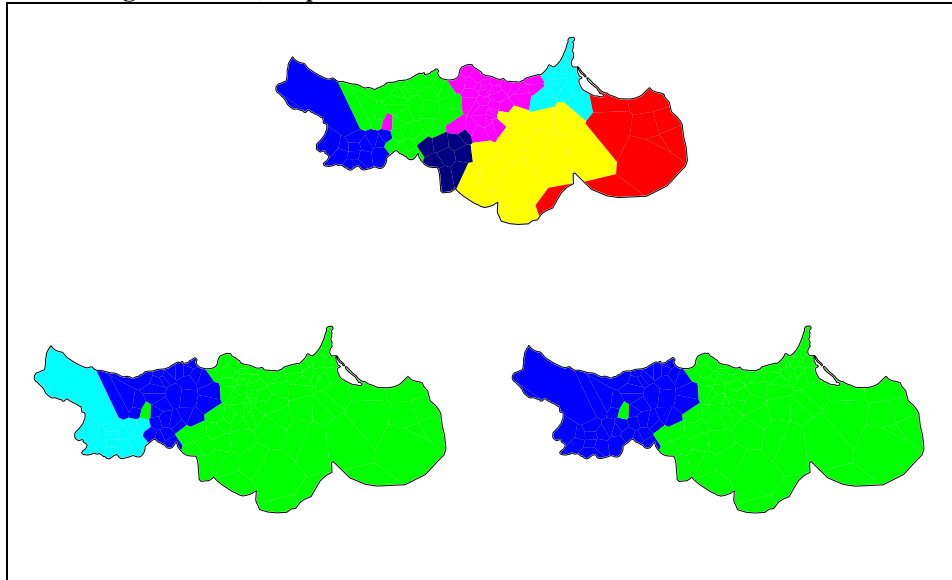


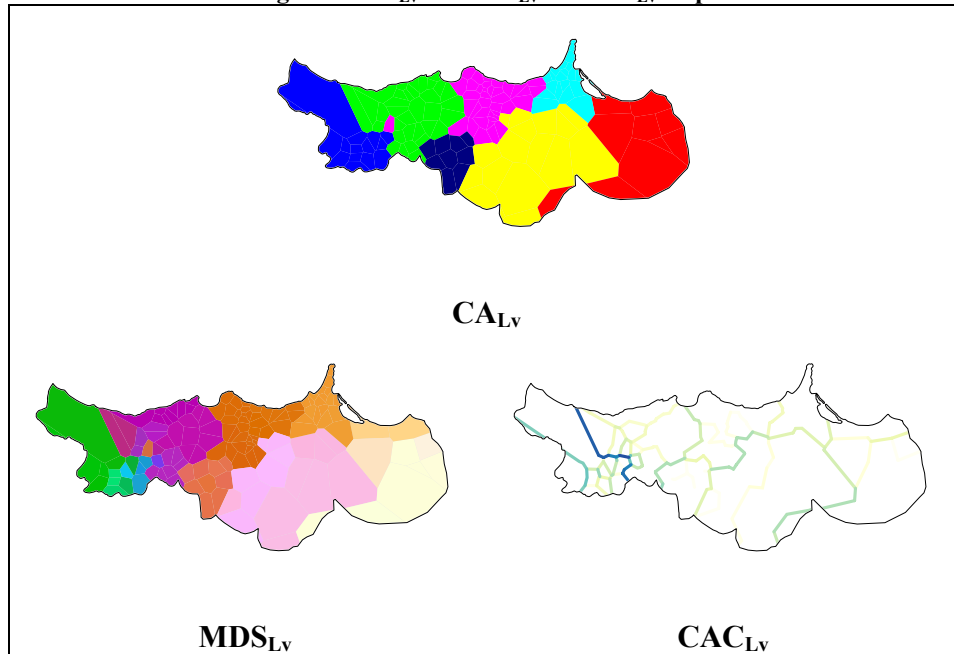
Figure 3: CA_{Lv} maps – 7 clusters vs. 3 clusters vs. 2 clusters – All lexis



3. Contrastive results

Because of its accuracy, the MDS method is most appropriate for dialectometry analysis of Berber lexis. Accordingly, it forms a yardstick against which other dialectometry methods can be contrasted (Figure 4). Among the Cluster Analyses classifications (CA), CA_{Bin} and CA_{Lv} join best the distribution maps displayed by MDS (7 groups). Moreover, the CA_{Lv} classification shows a further refinement because it takes into account the phonic variation of the units as much as their arrangement in the lexemes (Lafkioui 2008, 2009b). However, any analysis based on Lv distance (CA as well as MDS) ignores the existence of the hierarchy between the phonic units (phonetic units= phonemic units), unless various weights are granted to them through a specific parametring. This method implies the construction of a phonological system within the software, involving a time and energy-consuming effort that is much too expensive compared to its profits.

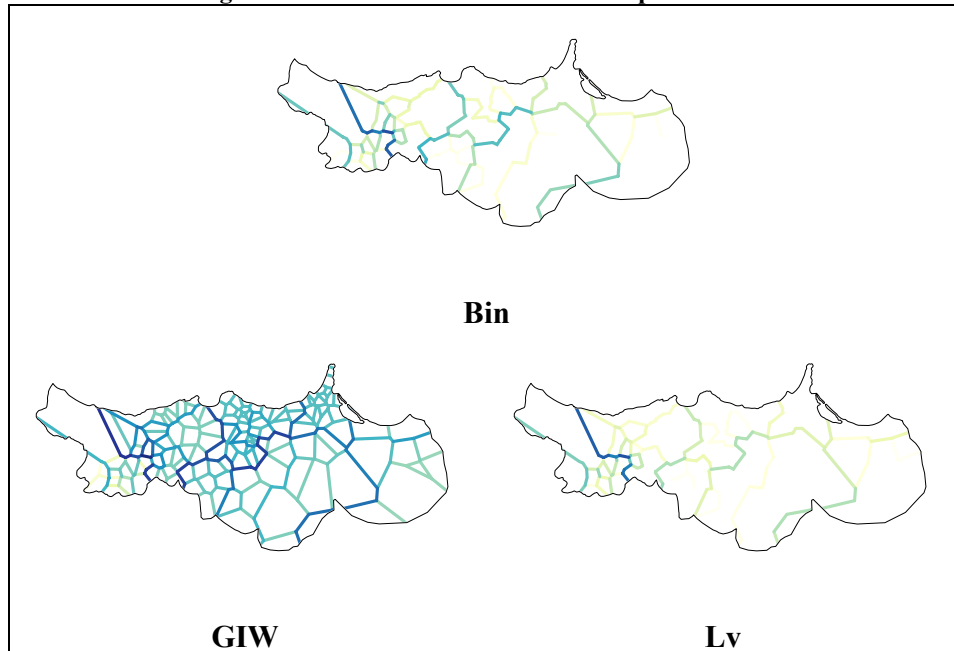
Figure 4: CA_{Lv} vs. MDS_{Lv} vs. CAC_{Lv} maps



The Cluster Analysis classification has the benefit of precisely indicating significant linguistic boundaries. The CAC maps (Composite Cluster map; Figure 5) designate these boundaries by dark lines. Compared to the distinctive boundaries drawn by the dendrogram and CA maps of Figures 3 and 4, the principal linguistic delimitation of the CAC map of Figure 5 is drawn further to the West. It is important, nevertheless, to note that the CAC maps do not seem suitable to display the classification of Rif-Berber lexis because of the difficulty of interpreting the data, due to their rather chaotic representation (Figure 5).¹⁰

¹⁰ Kleiweg (RuG/L04) offers some alternatives to the Ward algorithm which seems to be causing visual disorder by which CA_{GIW} maps are mostly affected.

Figure 5: CAC – Bin vs. GIW vs. Lv maps – All lexis



Dialectometry analysis, especially when based on the Levenshtein algorithm, is most valuable for the geographical classification of linguistic varieties, once the phonological system in question is integrated into the parametrizing of the software tool. This procedure is feasible on the level of languages (e.g. Tarifit) as well as language sub-families (e.g. Berber) and families (e.g. Afroasiatic) but is only productive – because it requires a great deal of time and energy – when applied to a considerable corpus concerning at least the sub-family level. An interesting development would be to build a common phonological parameter structure for each phylum, drawn from the micro-analysis of its various languages, which may lead to a general linguistic classification grid of Africa that allows comparing a substantial amount of features in a structured way, combining phonology, grammar and lexicon.

III. Linguistic geography

The term linguistic geography refers to the analysis of contact-induced language change and diffusion. It aims at delimiting linguistic zones geographically, on a micro- and macro-level, by reconstructing the long-standing affiliation patterns ensuing from non-genetic contact situations. Accordingly, a linguistic area is identified when the properties of a linguistic system specific to the region in question cannot be explained by genetic inheritance, universal principles, drift or chance. Well known examples of established linguistic areas, viz. linguistic convergence zones, are for instance Standard Average European (Haspelmath 2001) and the Mesoamerican linguistic area (Campbell et al. 1986).

From an African perspective, the key question that arises from this approach is the following: is Africa shaped areally?

The answer to this question is not only enlightening for Africa's linguistic classification – areally, (inter)genetically and typologically speaking – but also allows for a better understanding of its cultural, historical and ecological developments. Given the intricate and fused nature of these classification types, answering this question is a particularly tricky task.

However some seminal research on the areal classification of the languages of Africa has been carried out (see next paragraph), but there is still a lot of work ahead, both on a macro-level, considering Africa as one defined area, and on a micro-level, which is confined to a specific African sub-area and its sub-systems. The delayed interest in this research field is mainly associated with the lack of detailed comparative analyses of genetically related languages, which are essential to areal as well as intergenetic classifications.

One of the first scholars to take an interest in the study of African linguistic geography from a macro point of view was Greenberg (1959, 1983). He distinguished Africa's linguistic zones by means of the areal patterns of certain properties (mainly lexical polysemies, e.g. the employment of the same lexeme to express the notion of 'eat'/'defeat'/'have sexual intercourse' or to express the notion of 'meat'/'wild animal'), thus providing important accounts that have served as a basis for further studies (e.g. Heine & Nurse 2008). Although Greenberg's classification work has been criticised as unclear and inaccurate with regard to the distinction between areal and genetic parameters (Güldemann 2008), it has the merit of drawing attention to some of Africa's linguistic traits, the genetic origin of which he has also reconstructed.

Larochette's areal-typological contribution (1959) is perhaps less known than Greenberg's but is as important to the study of African linguistic geography. Most of his proposed features concern the languages of Congolese Bantu (Kicongo, Luba, Mongo), Mangbetu (Central Sudanic) and Zande (Ubangi), and are still employed to establish and expand other classifications (Meeussen 1975, Heine 1975, Heine & Leyew 2008).

Even if there is reason to believe that Africa constitutes (to a certain extent) a linguistic area, scholars have failed to find a significant cluster of features that are attested solely on African territory. However, a considerable effort to classify Africa, and more precisely sub-Saharan Africa, has been made by Heine (1975) and lately by Heine & Leyew (2008) on the basis of 11 typological features that have been the subject of a survey of 99 languages belonging to the four African phyla, namely the Afroasiatic phylum, the Nilo-Saharan phylum, the Niger-Congo phylum and the Khoisan phylum. Because of the poor results obtained from the isopleth mapping

method, which consists of delimiting areas of languages with the same number of features regardless of their nature (van der Auwera 1998: 260), Heine & Leyew considered typological resemblance on the basis of not only the presence of a feature but also its absence. Their findings corroborate the tendency to regard the Sudanic belt (West-Central Africa) as the area that contains most of Africa's linguistic characteristics, cutting across the boundaries of three genetic language families, i.e. the Afroasiatic phylum, the Nilo-Saharan phylum and the Niger-Congo phylum (Westermann 1911; Greenberg 1959, 1983). Fundamental properties distinguishing this major convergence area, such as the 'meat'/'animal' polysemy and labial-velar stops, indicate the Niger-Congo family as its nucleus and hence confirm Greenberg's (1983) reconstructions.

However, Güldeman's (2008) study of the Macro-Sudan belt, a northern sub-Saharan area extending from West Africa to Ethiopia in the East, disputes the validity of this macro-classification of Africa, pointing to the danger of carrying out areal research from a Sudanic-biased or even sub-Saharan-biased perspective.

Indeed, it is true that North and North-East Africa are typically excluded from this predisposed classification trend because of their dissimilar typological profile related to their common prevailing Afroasiatic genetic background. And yet these zones provide accounts of one of the indisputably typological linguistic features of Africa. This is the marked-nominative, a feature which is barely attested outside Africa but is present in most case languages within it, including those of Eastern Africa (König 2006, 2008). In these systems, it is the nominative that is functionally marked with reference to the accusative, even if the languages in question may differ as regards the morphosyntactic marking procedures, as shown in the following examples:

(a) Turkana (East Nilotic, Nilo-Saharan; Dimmendaal 1983: 263)

| | | |
|----------------------|-------------|-----------------|
| <i>è-sàk-ì</i> | <i>a-pà</i> | <i>a-k-ìmuḵ</i> |
| 3-want-ASP | father-NOM | food-ACC |
| 'Father wants food.' | | |

- The nominative is indicated by a floating low tone, whereas the accusative is the same as the basic form, also called the absolutive or absolute. All other cases are marked by fixed tonal schemes.

(b) Tamazight (Berber, Afroasiatic)

| | | |
|------------------------|---------------|-------------|
| <i>ɪ-swæ</i> | <i>ʊ-rgæz</i> | <i>æmæn</i> |
| he drank-PERF | man-NOM | water-ACC |
| 'The man drank water.' | | |

- The nominative is encoded by the morphological marker *ʊ-* (dependent state marker) and a continuous intonation unifying the noun with the preceding verbal syntagm, while the accusative is encoded by the absence of these morphemes and so matches the unmarked form; i.e. *æmæn* (accusative, independent state) versus *w-æmæn* (nominative, dependent state).

An exception to the abovementioned tendency to bias in the study of African linguistic geography is perhaps the contribution of Clements & Rialland (2008), who present a convincing areal classification of Africa as a whole that might be connected with contact-induced developments going back to the prehistory. They propose to divide Africa into six major zones – North, East, Sudanic, Center, South and Rift Valley – according to the presence or absence of the following specific phonological features: (1) absence of P-sounds, (2) emphatic consonants, (3) non-tonal prosody, (4) labial flaps, (5) labial-velar stops, (6) implosives, (7) nasal vowels, (8) two series of high vowels, (9) 3+ tone levels, (10) “lax” question markers, (11) ejective and aspirated stops, (12) clicks and (13) slack voiced stops. Although no single phonological characteristic common to the entire African continent is found, a certain areal regrouping is put forward, even if the dividing lines are considered permeable and unstable (Figure 6). From the six clusters, the first three display a clear areal demarcation cutting across genetic boundaries: the North zone is mainly defined by the properties (1), (2) and (3); the Sudanic belt is characterized by the properties (4), (5), (6), (7), (8), (9) and (10); the South is well distinguished by the properties (11), (12) and (13). The other three areal zones are less well delimited: apart from a shared Afroasiatic heritage, the East zone shows a prevalent presence of implosives, ejectives and tone systems; besides the common Bantu properties, the Center zone shares the vowel harmony system; as to the complex Rift Valley zone, no conclusive phonological evidence diffused over the entire area and so intersecting the genetic phyla is attested, which definitely calls for further study.¹¹

Figure 6: The six phonological zones of Africa (Clements & Rialland 2008)



¹¹ For more about the Rift Valley area, see Kiessling et al. (2008).

On a mid- and micro-level, a considerable amount of research on African areal classification has been accomplished, although not always backed up with conclusive data. The Ethio-Eritrean area – also called the Ethiopian area – is considered the most defined, if not the only, convergence zone of Africa (e.g. Greenberg 1959; Ferguson 1970, 1976; Heine 1975; Appleyard 1989, Zaborski 1991, 2003; Hayward 2000; Crass 2002; Bender 2003). This area is demarcated and expanded differently according to the traits used for analysis. Heine's (1975) constituent order classification, for example, enlarges the area by including the Lake Chad, while Tosco's (1996) interest in the study of the verb paradigms limits it to the Gurage-Sidamo terrain. With the exception of Tosco (2000), the ongoing debate generally supports the hypothesis of the Ethio-Eritrean convergence area, which has been recently extended to the entire African horn and certain parts of Sudan, Tanzania and Uganda (Zaborski's 2003).

IV. Conclusion

Language is continuously modulated through innovations that may surface in structurally layered and causal formations dictated by properties that are system-internal and/or system-external. It goes without saying that social factors, such as prestige and community solidarity, are important for the diffusion of variants, which may follow different adoption and acquisition scenarios. However, the Berber data concerning the phonological and morphological phenomena produced by vocalisation of the liquids /r/ and /r̥/ in Tarifit (Berber of North Morocco) and examined in Lafkioui (2009c, 2011b) reveal the importance of certain system-internal factors in the diffusion of new linguistic items. For example, economy and code conformity are important functional parameters for the dispersal of the vocalised liquids in the Central Rif area. Moreover, structure adequacy and systematisation/generalisation of specific morphological patterns play a significant role in the success of their diffusion across the Tarifit varieties. Contrary to the language evolutionary claim that only social factors are responsible for variant selection (Milroy & Milroy 1985; Milroy 1992: 201-202; Croft 2000: 38, 39, 54), Lafkioui demonstrates how functional and social (socioeconomic) factors can interact in the selection and hence diffusion of language forms and how in some cases, as in the restructuring of the verbal paradigm, system-internal properties may prevail.

Another interesting finding related to linguistic variational research is that language change is gradual not just on an extra-linguistic level (geographical and social variation) but also on a linguistic one. Therefore, it is important to consider the continuous selection process of variants not exclusively from a social perspective ("propagation" of Croft 2000: 38, 178) but also in terms of how the variants are formally and functionally incorporated into ever-changing linguistic structures (Lafkioui 2009c, 2011b).

It should also be noted that there exists an important correlation between linguistic variation and ecology (besides physical geography and technology). The ecological risk hypothesis states that in non-industrial societies, such as certain West African ones (Nettle 1996), social networks serve as a kind of protection against the risks of agriculture (Scott 1976; Braun & Plog 1982).¹² In this approach, geolinguistic diversity is employed as historical evidence to explain the particular

¹² The bigger the provision problem, the larger the social network.

Mena Lafkioui, *Dialectology and linguistic geography*, in: Vossen (ed.), *Handbook of African Languages*, Oxford: Oxford University Press, forthcoming in 2012.

social and economic networks formed in order to adapt to the ecological situation at hand and so reduce the risk of undernourishment. It would be of great interest to consider things the other way around; that is, to study the effects of ecology on language, considering linguistic diversity also as an outcome of people's conduct rather than just a consequence of physical geography. Unfortunately, this field of study is by and large lacking in African geolinguistics.¹³

V. References

- Appleyard, D. 1989. The relative verb in focus construction: an Ethiopian areal feature. *Journal of Semitic Studies*, 34 (2): 291-305.
- Auwers, J. van der 1998. Revisiting the Balkan and Meso-American linguistic areas. *Language Sciences*, 20 (3): 259-270.
- Basset, A. 1929. *Etudes de géographie linguistique en Kabylie*. Paris: Leroux.
- Basset, A. 1959. *Articles de dialectologie berbère*. Paris: Klincksieck.
- Bastin, Y. 1983. Classification lexicostatistique des langues bantoues (214 relevés). *Le Bulletin des Séances de l'Académie Royale des Sciences d'Outre-Mer*, 27: 173-199.
- Bender, M. L. 2003. Northeast Africa: a case study in genetic and areal linguistics. *Annual Publication in African Linguistics*, 1: 21-45.
- Braun D. P. & Plog S. 1982. Evolution of tribal social networks: Theory and prehistoric North American evidence. *American Antiquity*, 47: 504-525.
- Brunk, K. et al. 1999. L'atlas linguistique d'Afrique sahélo-soudanienne (Alass). À la recherche des isoglosses intergénétiques dans le domaine zoonymique. In: Baroin & Boutrais (eds.), 171-179.
- Campbell L. et al. 1986. Meso-America as a linguistic area. *Language*, 62 (3): 530-570.
- Chambers, J. K. & Trudgill, P. 1980. *Dialectology*. Cambridge: Cambridge University Press.
- Clements, N. & Rialland, A. 2008. Africa as a phonological area. In: Heine & Nurse (eds.), 36-87.
- Crass, J. 2002. Ejectives and pharyngeal fricatives: two features of the Ethiopian language area. In: Yimam et al. (eds.), 1679-1691.
- Croft, W. 2000. *Explaining Language Change: An evolutionary approach*. Harlow: Longman.
- Daan, J. C. 1999. Dialects. In: Preston (ed.), 9-30.
- Daan, J. C. & Blok, D. P. 1969. *Van Randstad tot Landrand; toelichting bij de kaart: Dialecten en Naamkunde*. Amsterdam: Noord Hollandsche Uitgevers Maatschappij.
- Dimmendaal, G. J. 1983. *The Turkana Language*. Dordrecht: Foris Publications.
- Ferguson, Ch. A. 1970. The Ethiopian language area. *Journal of Ethiopian Studies*, 8 (2): 67-80.
- Ferguson, Ch. A. 1976. The Ethiopian language area. In: Bender et al. (eds.), 63-76.
- Forquet, J. 1956. Linguistique structurale et dialectologie. *Fragen und Forschungen im Bereich un Umkreis der germanischen Philologie: Festgabe für Theodor Frings zum 70. Geburtstag*, Berlin: Akademie-Verlag, 190-203.
- Garde, P. 1961. Réflexions sur les différences phonétiques entre les langues slaves. *Word*, XVII: 34-62.
- Gilman, Ch. 1986. African areal characteristics: sprachbund, not substrate? *Journal of Pidgin and Creole Languages*, 1(1): 33-50.

¹³ Except perhaps for Brunk et al. (1999).

Mena Lafkioui, *Dialectology and linguistic geography*, in: Vossen (ed.), *Handbook of African Languages*, Oxford: Oxford University Press, forthcoming in 2012.

- Goebel, H. 1982. *Dialektometrie. Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie*. Vienne: Verlag der österreichischen Akademie der Wissenschaften.
- Goebel, H. 1993. Probleme und Methoden der Dialektometrie: Geolinguistik in globaler Perspektive. In: Viereck (ed.), 37-81.
- Gooskens, C. 1997. *On the Role of Prosodic and Verbal Information in the Perception of Dutch and English Varieties*. PhD. Dissertation. Nijmegen: Katholieke Universiteit Nijmegen.
- Goossens, J. 1965. *Die niederländische Strukturgeographie und die "Reeks Nederlandse Dialectatlassen"*. Amsterdam: Noord-Hollandsche Uitgevers Maatschappij.
- Goossens, J. 1969. *Strukturelle Sprachgeographie. Eine Einführung in Methodik und Ergebnisse*. Heidelberg: Winter.
- Goossens, J. 1977. *Inleiding tot de Nederlandse Dialectologie*. Groningen: Wolters-Noordhoff.
- Greenberg, J. H. 1959. Africa as a linguistic area. In: Bascom & Herskovitz (eds.), 15-27.
- Greenberg, J. H. 1983. Some areal characteristics of African languages. In: Dihoff (ed.), 3-22.
- Grosse, R. 1960. Strukturalismus und Dialektgeographie. *Biuletyn Fonograficzny*, III: 89-101.
- Güldemann, T. 2008. The Macro-Sudan belt: towards identifying a linguistic area in northern sub-Saharan Africa. In: Heine & Nurse (eds.), 151-185.
- Guthrie, M. 1967-71. *Comparative Bantu: an Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg International.
- Haspelmath, M. 2001. The European linguistic area: Standard Average European. In: Haspelmath et al. (eds.), 1492-1510.
- Hayward, R. J. 2000. Is there a metric for convergence? In: Renfrew et al. (eds.), 621-640.
- Heeringa, W. 2004. *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. PhD. Dissertation. Groningen: Rijksuniversiteit Groningen.
- Heine, B. 1971. Kalenjin Glottochronology. Preliminary hypotheses. *Mila*, 2 (2): 51-55.
- Heine, B. 1974. Historical linguistics and lexicostatistics in Africa. *Journal of African Languages*, 11 (3): 7-20.
- Heine, B. 1975. Language typology and convergence areas in Africa. *Linguistics*, 144: 27-47.
- Heine, B. & Leyew, Z. 2008. Is Africa a linguistic area? In: Heine & Nurse (eds.), 15-35.
- Heine, B. & Möhlig, W. J. G. 1980. *Language and Dialect Atlas of Kenya. Volume I: Geographical and Historical Introduction*. Berlin: D. Reimer.
- Heine, B. & Nurse D. (eds.) 2008. *A linguistic geography of Africa*. Cambridge: Cambridge University Press.
- Hoppenbrouwers, C. & Hoppenbrouwers, G. 1988. De feature frequentie-methode en de classificatie van Nederlandse dialecten. *Bulletin voor taalwetenschap*, 18(2): 51-92.
- Hoppenbrouwers, C. & Hoppenbrouwers, G. 2001. *De indeling van de Nederlandse streektaalen: Dialecten van 156 steden en dorpen geklasseerd volgens de FFM (feature frequentie methode)*. Assen: Koninklijke Van Gorcum.
- Kessler, B. 1995. Computational dialectology in Irish Gaelic. *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*. San Francisco: Morgan Kaufmann, 60-67.
- Kiessling R. et al. 2008. The Tanzanian Rift Valley area. In: Heine & Nurse (eds.), 186-227.

Mena Lafkioui, *Dialectology and linguistic geography*, in: Vossen (ed.), *Handbook of African Languages*, Oxford: Oxford University Press, forthcoming in 2012.

- Kocks, G. H. 1970. *Die Dialekte von Südostdrente und anliegenden Gebieten. Eine strukturgeographische Untersuchung*. PhD. Dissertation. Groningen: Rijksuniversiteit Groningen.
- König, C. 2006. Marked nominative in Africa. *Studies in Language*, 30 (4): 705-782.
- König, C. 2008. The marked-nominative languages of eastern Africa. In: Heine & Nurse (eds.), 251-271.
- Kruskal, J. 1999. An overview of sequence comparison. In: Sankoff & Kruskal (eds.), 1-44.
- Lafkioui, M. 2007. *Atlas linguistique des variétés berbères du Rif*, Köln, Rüdiger Köppe Verlag.
- Lafkioui, M. 2008. Dialectometry Analyses of Berber Lexis. *Folia Orientalia*, 44: 71-88.
- Lafkioui, M. 2009a. Etudes de géographie linguistique berbère: variation géolinguistique et classification dialectométrique, *Le Bulletin des Séances de l'Académie des Sciences d'Outre-Mer*, 54(4): 439-460.
- Lafkioui, M. 2009b. Analyses dialectométriques du lexique berbère du Rif. In : Ibriszimow & al. (eds.), 133-150.
- Lafkioui, M. 2009c. Synchronic and diachronic linguistic variation as an indicator of language change and diffusion in Tarifit. *Studi Magribini* (Nuova Serie), 7: 65-84.
- Lafkioui, M. 2011a. Variation et structuration linguistiques et sociolinguistiques en berbère du Rif. Köln: Rüdiger Köppe Verlag.
- Lafkioui, M. 2011b. How system-internal linguistic factors indicate language change and diffusion. A geolinguistic analysis of Berber data. *Dialectologia et Geolinguistica*, 19: 62-80.
- Larochette, J. 1959. Overeenkomst tussen Mangbetu, Zande en bantu-talen. *Handelingen van het XXIIIe Vlaams Filologencongres*, Brussel, 247-248.
- Martinet, A. 1972. Structural dialectology. *Pakha Sanjam*, 5 (Special volume), article n° 2.
- Meeussen, A. E. 1975. Possible linguistic Africanisms. *Fifth Hans Wolff Memorial Lecture*. Bloomington: Indiana University.
- Milroy, J. 1992. *Linguistic Variation and Change*. Oxford: Blackwell.
- Milroy, J. & Milroy, L. 1985. Linguistic change, social network, and speaker innovation. *Journal of Linguistics*, 21: 339-384.
- Möhlig, W. J. G. 1974. *Die Stellung der Bergdialekte im Osten des Mt. Kenya. Ein Beitrag zur Sprachgliederung im Bantu*. Berlin: D. Reimer.
- Möhlig, W. J. G. 1979. Dialektgrenzen und Dialektcontinua im Bantu-Sprachgebiet von Kenia. Zum Problem der Grenzfindung und Grenzgewichtung. *Tagungsband des Internationalen Symposions "Zur Theorie des Dialekts"*, Marburg 1977. Wiesbaden: Franz Steiner Verlag.
- Möhlig, W. J. G. 1980. L'atlas linguistique du Kenya: méthodes et résultats. In: Bouquiaux et al. (eds.), 57-83.
- Moulton, W. G. 1960. The short vowel systems of northern Switzerland. *Word*, 16: 155-182.
- Nerbonne, J. & Heeringa, W. 1998. Computationale vergelijking en classificatie van dialecten. *Taal en Tongval*, 50(2): 164-193.
- Nerbonne, J. & Heeringa, W. 2001. Computational Comparison and Classification of Dialects. *Dialectologia et Geolinguistica*, 9: 69-83.
- Nettle, D. 1996. Language Diversity in West Africa: An Ecological Approach. *Journal of Anthropological Archaeology*, 15: 403-438.
- Rensink, W. G. 1955. Dialectindeling naar opgaven van medewerkers. *Mededelingen der Centrale Commissie voor Onderzoek van het Nederlandse Volkseigen*, 7: 20-23.

Mena Lafkioui, *Dialectology and linguistic geography*, in: Vossen (ed.), *Handbook of African Languages*, Oxford: Oxford University Press, forthcoming in 2012.

- Salikoko, M. S. 2000. From Genetic Creolistics to Historical Dialectology: Ecological and Population Genetics Perspectives. *American Speech*, 75 (3): 262-265.
- Saussure, F. de 1973 [1916]. *Cours de linguistique générale*. T. de Mauro (ed.). Paris: Payothèque.
- Scott, E. P. 1976. *Indigenous Systems of Exchange and Decision-making among Small-holders in Rural Hausaland*. Ann Arbor: Department of Geography Publications.
- Séguy, J. 1973. La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de linguistique romane*, 37: 1-24.
- Sommer, G. & Vossen, R. 1993. Dialects, sectiolects, or simply lects? The Maa language in time perspective. In: Spear & Waller (eds.), 25-37.
- Stankiewicz, E. 1957. On discreteness and Continuity in Structural dialectology, *Word*, 13: 44-59.
- Tosco M. 1996. The Northern Highland East Cushitic verb in an areal perspective. In: Griefenow-Mewis and Voigt (eds.), 71-100.
- Tosco, M. 2000. Is there an "Ethiopian language area"? *Anthropological Linguistics*, 42 (3): 329-365.
- Vossen, R. 1988. *Towards a Comparative Study of the Maa Dialects of Kenya and Tanzania*. Hamburg: Helmut Buske.
- Vossen, R. 2008. How to comprehend the obvious? Linguistic variation and dialect geography: the African experience. *Sprache und Geschichte in Afrika*, 19: 201-224.
- Weijnen, A. (1946). De grenzen tussen de oost-noord-Brabantse dialecten onderling. *Bijdragen en mededelingen der Dialectencommissie van de Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam*, VIII: 1-15.
- Weijnen, A. 1966. *Nederlandse dialectkunde*. Assen: Van Gorcum.
- Weinreich, U. 1954. Is a structural dialectology possible? *WORD*, 10: 388-400.
- Westermann, D. H. 1911. *Die Sudansprachen*. Hamburg: Friederichsen.
- Zaborski, A. 1991. Ethiopian language subareas. In: Pilaszewicz & Rzewuski (eds.), 123-134.
- Zaborski, A. 2003. Ethiopian language macroarea. *Sprawozdania z Posiedze ń Komisji Naukowych Oddziału PAN w Krakowie*, 45 (2): 60-64.