Linguistic Atlas of Turkish Azari in North West of Iran

Hiwa Asadpour [1], Faryar Akhlaghi [2]

[1.] Department of Linguistics, Science and Research branch, Islamic Azad University, Tehran, Iran
[2.] Languages and Dialects Research Center, Faculty of linguistics, Inscriptions, and Texts; Research Institute to Iranian Cultural Heritage and Tourism
Hiva.dandelion@gmail.com, faryar.a@gmail.com

Abstract: In this paper a range of methods for measuring the phonetic distance between dialectal variants are described. It concerns variants of methods as wordnet method and graded map analysis of linguistic levels. In addition, all features like simple (based on atomic characters) and complex (based on feature bundles) have been studied. The dialects were compared with each other directly and indirectly via a standard dialect. The results of comparison were classified by clustering and by training of a multidimensional map. The results were compared to well established scholarship in dialectology, yielding a calibration of the methods like information visualization technique. These results indicate that computational techniques are more sensitive in feature representations of dialects and such visualizations of information have good measures of phonetic overlap of feature bundles. The results of clustering give the sharper classification, but the graded map is a nice supplement. The findings show that Azari has composed of different regional groups which are relate to one ancestor which it might be the proto-Turkic language and it is not a group of languages.

Keywords: Linguistic Atlas, Computational Linguistics Language, Dialect, Kurdish, Azerbaijan-e Qarbi

1. Introduction

In the last quarter of 19th B.C. the study of language change led to an interest in dialects, and a number of scholars in Germany, Switzerland, France and Italy began to investigate regional variation in language. The oldest branch of dialectology is the study of what is today often referred to as "dialect geography", i.e. the study of the geographical distribution of language varieties (Falk cited in Asadpour, 2007: 50-52; Breton, 1991), as opposed to the study of many other relations between language varieties and external conditioning factors, such as social class, gender, age, identity, religion, occupation, economic status, education, and ethnicity (Labov, 1990, 1996; Milroy & Milroy, 1992; Kerswill, 2004). While it is clear that geography has a massive influence on the distribution of language varieties, and that closer varieties are normally more linguistically alike than more distant ones, still there have been surprisingly few attempts to examine these relationships with an eye toward more general Formulations (Chambers, 1995; Trudgill, 1980, 1986; Chambers et al., 2002).

Traditional dialectology relies on identifying language features which are common to one dialect area while distinguishing it from others. It has difficulty in dealing with partial matches of features and with non-overlapping language patterns. Attempts to delimit a dialect by topographical, political, or administrative boundaries ignore the obvious fact that within any such boundaries there will be variation for some features, while other variants will cross the borders. Similar oversimplification arises from those purely linguistic definitions that adopt a single feature to characterize a large regional complex. A dialect atlas in fact displays a continuum of overlapping distributions in which the "isoglosses" delimiting dialectal features vary from map to map and "the areal transition between one dialect type and another is graded, not discrete" (Benskin cited in Margaret & Lass, 2006).

The primary tool of traditional dialectology has been the isogloss, the delineation of a concrete language variation on a map. Language variants distinguished by many isoglosses emerge then as relatively distinct dialects. But dialectologists recognize that the method of isoglosses does not result in the delineation of "dialects" satisfactorily so we applied some new approaches as dialectometry. Dialectometry provides the more general tools with which such relationships may be studied (Goebl, 1982, 1984), and the present paper is an attempt to apply dialectometry to evaluate Trudgill's ideas more systematically. In fact it has been common to examine the dependence of dialect distance on geography from the earliest work on in dialectometry (Séguy, 1971; Heeringa & Nerbonne, 2001; Gooskens & Heeringa, 2004). There has been no systematic examination of Trudgill's gravity hypothesis from a dialectometric perspective, however. Dialectometry produces

aggregate distance matrices in which a distance is specified for each pair of sites. By projecting groups obtained by clustering onto geography one compares results with traditional dialectology, which produced maps partitioned into implicitly non-overlapping dialect areas. The importance of dialect areas has been challenged by proponents of continua, but they need too effort to compare their findings to older literature, expressed in terms of areas.

Azari is believed to have been a part of the dialect continuum of Northwest Iranian languages. It is also spoken by Turkish people in the Eastern part of south and north of Western Azerbaijan. Azari was spoken in Azerbaijan at least up to the 17th century, with the number of speakers decreasing since the 11th century due to the Turkification of the area. As such, its ancestor would be close to the earliest attested Northwest Iranian languages, Median. Some of Turkish people migrated in past centuries to the North of Khorasan.

More specifically, there is a dearth of in-depth quantitative and qualitative dialectology research both inside and outwith Iran concentrating specifically on geographical and social evaluations of varieties of Azari, as the very limited numbers of previous studies conducted amongst Iranian scholars have either been qualitative in design or too small in scale. Moreover, the findings of these studies have been somewhat inconclusive. In a quantitative and to somewhat qualitative study, we want to employ a range of innovative direct and indirect techniques of attitude measurement, investigated the varieties of Iranian Azari speakers and we intend to suggest linguistic groupings of the local speech varieties, their boundaries, their relation to each other and to other languages in this province as well as an approximation of the number of speakers for each variety. This research is the subset of a larger set included in the Linguistic Atlas of Western Azerbaijan (Asadpour, 2007) and its aims is to contribute to the wider goal of researching all the speech varieties in the Islamic Republic of Iran.

2. The Azari dialect atlas and its Objectives

The overall aim of this research is to investigate societal and geographical distributions of 300 linguistic items which involved phonological, lexical, morphological and syntactical variations across Azarbaijan-e Qærbi. From this point, Azarbaijan-e Qærbi is the only province in Iran that it has common geographical borders with four neighboring countries, Turkey, Azerbaijan, Iraq and Armenia, it is also one of the commercial regions of Iran, and five languages of Persian, Azari, Kurdish, Armenian, and Assyrian are spoken there. The second goal of this paper is to contribute to the understanding of the geographic

distribution of linguistic variation, and to argue that these distributions reflect the dynamics of linguistic diffusion. This argumentation effectively uses (aggregate) synchronic distributions as evidence of diachronic patterns of diffusion.

So, we compared string distance measures for their value in modeling dialect distances. Traditional dialectology relies on identifying language features which are common to one dialect area while distinguishing it from others. It has difficulty in dealing with partial matches of linguistic features and with non-overlapping language patterns. Therefore Séguy (1973) and Goebl (1982; 1984), advocate using aggregates of linguistic features to analyze dialectal patterns, effectively introducing the perspective of dialectometry. Although we wish to contribute to the understanding of the general principles underlying the geographic distribution of linguistic variation (Orton etal., 1978; Petyt, 1980), we structure our paper as a test of the very specific aggregate hypothesis, according it the attention we feel it deserves as an early attempt at a general formulation of the principles of how geography influences variation.

3. Data collection

The current paper is based on findings of a larger project the Linguistic Atlas of Azarbaijane Qarbi (LAAQ) as an MA thesis by Asadpour (2011). The project was based on different approaches and different tools like typological questioanre which has been designed on the foundation of Comrie (1977) and some other field instruments in Max-Planck website. Both elicited data and non-elicited data have been gathered for further studies and the voice of informants has been recorded by the researcher[1].

4. Computational and Visualization Methods as New Analyzers in Dialect Detection

Jean Séguy was director of the *Atlas linguistique de la Gascogne*. He and his associates published six atlas volumes. In these volumes maps are published in which single answers were plotted (Chambers & Trudgill, 1998: 137). However, Séguy looked for a way to analyze the maps in a more objective way than was possible with traditional analytic methods. For each pair of contiguous sites Séguy and his research team counted "the number of items on which the neighbors *disagreed*." The number of disagreements between two neighbors was expressed as a percentage, "and the percentage was treated as an index score indicating the *linguistic distance* between any two places" (Chambers & Trudgill, 1998: 138, see also Séguy, 1971 & 1973).

---

[1]- for more information on data collection and methodology see Asadpour (2011)

The items fell into five types: 170 lexical variables, 67 pronunciations, 75 phonetic/phonological, 45 morphological and 68 syntactic. Séguy weighted all types equally by calculating percentages for each type rather than for each item. The final linguistic distance was calculated as the mean of the five percentages.
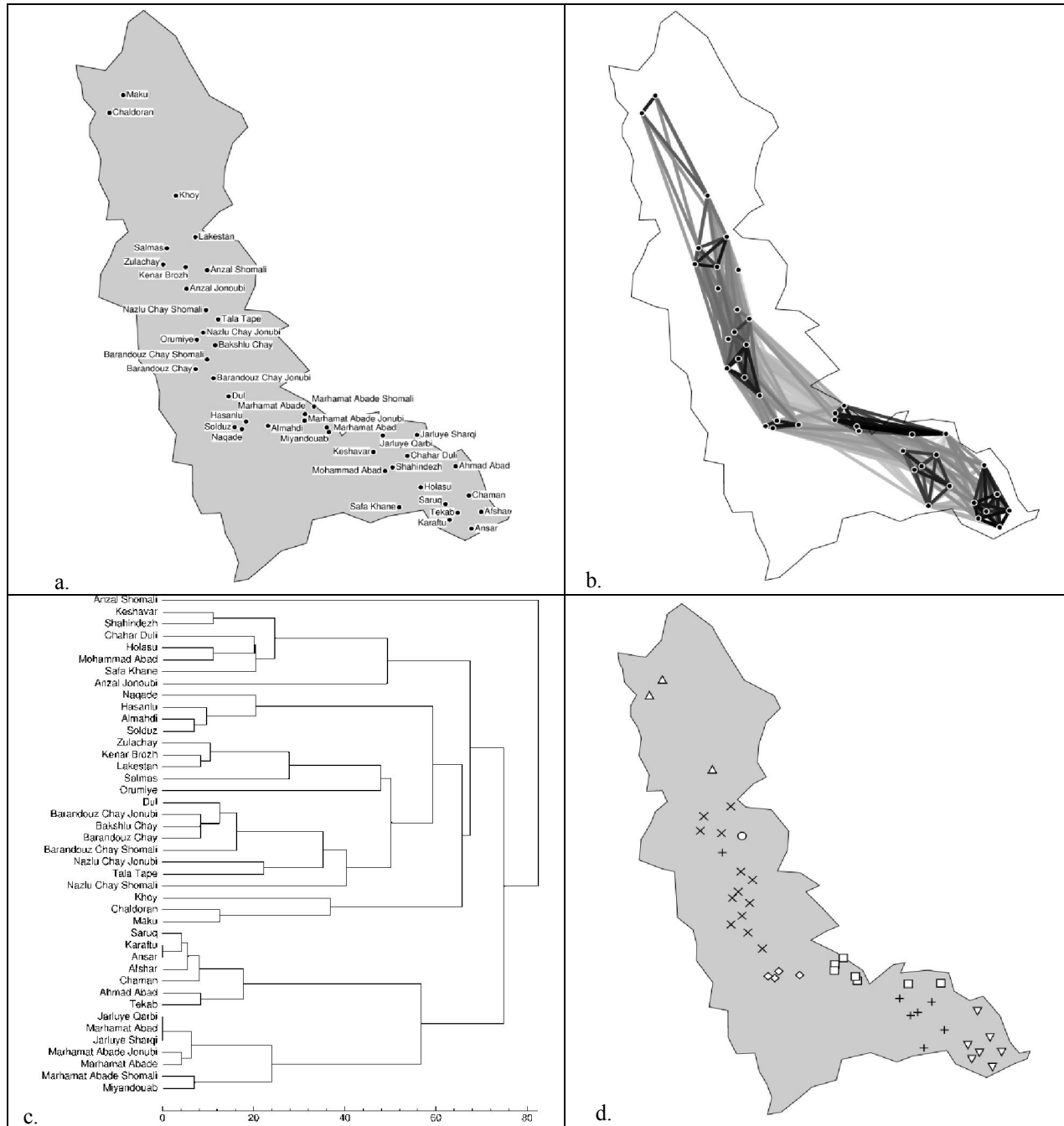

Figure 1. Figure a. shows the locations of the 42 Azerbaijan dialect varieties. Figure b. shows the lexical distances among them. Darker lines represent small distances, lighter lines represent larger ones. With cluster analysis we obtained the dendrogram shown in Figure c. The seven most significant groups are shown in Figure d.

Séguy and his team calculated the linguistic distances for each item, for each item type and for the composites. They were plotted on maps, which can be found in the last ten pages of the sixth volume of the atlas which was published in 1973. Strongly related to the methodology of Séguy is the work of Goebl,

although the basis of Goebl's work was developed mainly independent of Séguy (see Goebl, 1982 en Goebl, 1993).

Just as Séguy we want to analyze the newly collected data of the *Linguistic Atlas of Azerbaijan-e Qarbi* in an objective way in order to find the main patterns suggested by the atlas data. we follow Séguy and calculate the distance between two dialect varieties as the number of items on which they disagree divided by the total number of items which is considered. But here in this study we calculate distances based on lexical and phonological levels. The number of items on which the two varieties disagree is divided by 100 linguistic items. we analyze the distances further with hierarchical cluster analysis. The goal of clustering is to identify the main groups. The groups are called *clusters*. Clusters may consist of sub-clusters, and sub-clusters may in turn consist of sub-sub-clusters, etc. The result is a hierarchically structured tree in which the dialects are the leaves (Jain and Dubes, 1988). Later on as Historical linguists we will design a newly wordnet method for comparison of linguistic levels. The result is strong correlations among linguistics levels and items.

On the basis of distances among the dialect varieties, the varieties are clustered with cluster analysis. The result is a dendrogram for each level. The dendrograms are shown in Figures 1c, 2c and 3c respectively. The seven most significant groups as suggested by the dendrograms are shown in Figures 1d, 2d and 3d respectively. The classifications of these Figures correspond with the line maps for the greater part, but also reveal some details which cannot be clearly seen in the line maps. At the lexical level the cluster map suggests six groups, where the variety of Anzal Shomali does not belong to any group. It is also striking that the variety of Anzal Janoubi belongs to the geographically distant group with the varieties of Saruq, Karaftu, Ansar, Afshar, Chaman, Ahmad Abad and Tekab. The morphological cluster map is quite similar to the lexical one, with again the exceptional position of Anzal Shomali. The variety of Anzal Jonoubi now belongs to the group with the geographically close varieties of Nazlu Chai Shomali, Tala Tape, and others. At the syntactic level the three northern groups in the line map form one group in the cluster map. But several varieties which are found in these groups geographically do not belong to this group. The varieties of Tala Tape and Nazlu Chay Jonoubi belong to the southern group with the varieties of Saruq, Chaman, and others. Anzal Shomali – at the lexical level already to be found exceptional – is clustered together with the varieties of Solduz, Almahdi and others. The varieties of Anzal Shomali and Anzal Jonoubi do not belong to any

groups, but some relationship between the varieties is suggested by the dendrogram.

Continuum map (see the correlations in the appendix)
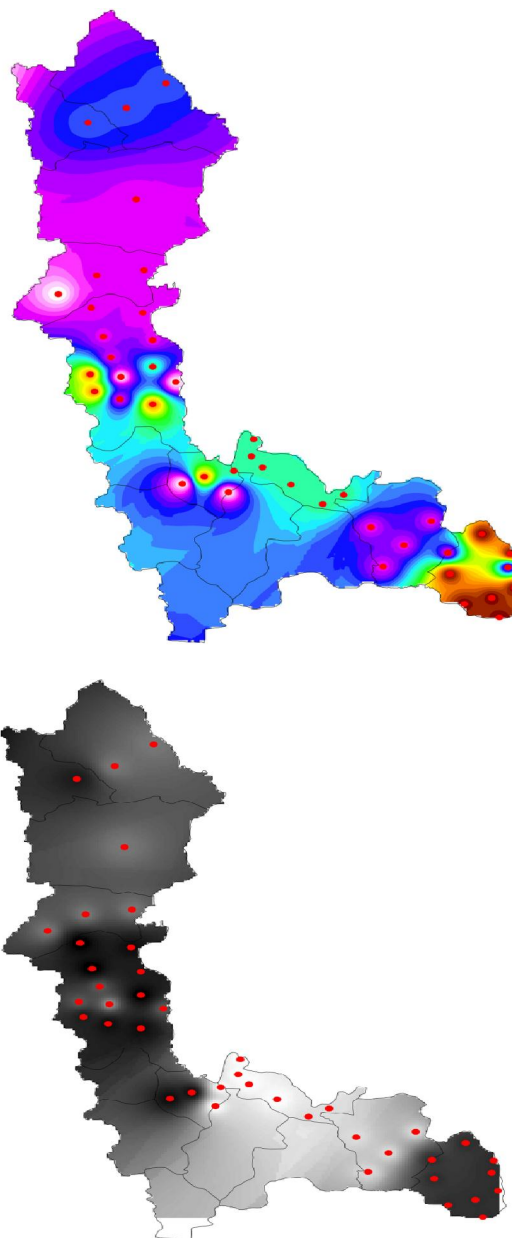
The map in bottom is the continuum black and white map for jump down. The generated colorful map for father is based on lexical similarities and the other colorful map for knee is based phonological similarities based on calculation of differences among dialects. As it is shown on the maps there are isolated areas which we ignore them but as for the rest of the maps there is a unity among dialects. This unity and

continuum is best shown on black and white map. It is clear that due to topographic situation of the regions some changes occur among Azari dialects and many extralinguistic factors like economy, history, language contact with neighboring languages which are in Spraschbund as kurdish, Armenian, Assyrian-NeoAramaic and Persian but these are not the focus of this study. What is important the regular changes from south to north. All statistical analysis and visualizations present that there is no language border in reality and it is the matter of geography that made some sharp changes in language continuum.

5. Conclusions and Prospects

Based on our analysis, our data show that among all levels in Azarbaijan-e Qærbi, phonological features seem to be more sensitive to regional and social differences than morphological and syntactic features, so we make a distinction between accent and dialect that in Azari we have different sub-accents rather than dialects. In addition, the distinction between standard and non-standard accent in Azari communities is not clear and there is no prestigious accent. There may be a general difference between items of pronunciation and other items (morphology, syntax and lexicology), in that pronunciation is less liable to standardization. Given, the special connection between standardization and writing, it would not be surprising if this were so. In Azarbaijan-e Qærbi pronunciation seems to have a different social function from other types of items and despite the apparent influence of Persian and Kurdish, its influence is restricted almost entirely to vocabulary and appears to have had no affect at all on the pronunciation of even the most susceptible groups. Azari speakers use pronunciation perceptions in order to identify their origins so for them their region and origin is very important. But the differences which

exist in other levels could be an artifact of the process of language planning by the Iranian Academy of Persian Language and Literature.

Existing literature, knowledgeable informants and the perception of local speakers concur that Azari varieties are the same. Taking into account the results of lexico- and phonostatistic comparison the most genuine conclusion would be that all speech forms in the Azarbaijan-e Qærbi originate from Azari (or a proto-language that Turkic languages is similar to) and are closely related to varieties in Azerbaijan and Turkey. Taking into account people's perception, the number of speakers and the existing infrastructure of the Azarbaijan-e Qærbi it becomes obvious that Orumiye (city) has to play a major role. Considering the fact that the lexico-, morpho-, syntactico- and phonostatistic similarity of Azari varieties are quite high. We decided that it is most natural to propose that there are different sub-accents in the region besides other languages in the province, which according to literature, knowledgeable informants, people's perception and lexico-, morpho-, and syntactico- as well as phonostatistic comparison is part of the Western Iranian language group. Another way to view the linguistic situation is to divide the region into two axes. One from the North to the Center, coming from Maku and one from the Center to the South. The second one, much more homogeneous, is spread out in the plain between the mountain chain and the desert. The first axis shows great linguistic diversity, which is not surprising, as it goes through the mountains.

Finally, it would be interesting to apply these techniques to situations in which language genealogies are at issue, but this would seem to require some means of excluding similarities due to borrowing.

Appendix

Lexical correlation coefficient

	father	raw	sun	a kind of bird which like snow	smoke	nail	Cousin	tree	light	fist	afternoon	leg
father	1	,334[*]	,144	,356[*]	,108	,272	,189	-,146	,336[*]	,092	,105	-,286
raw	,334[*]	1	-,034	-,080	,208	,415[**]	-,075	,034	,278	,079	,407[**]	-,198
sun	,144	-,034	1	-,295	-,070	,212	-,072	-,291	,272	,000	,079	-,648[**]
a kind of bird which like snow	,356[*]	-,080	-,295	1	,047	-,070	,054	,210	-,251	-,189	-,157	,473[**]
smoke	,108	,208	-,070	,047	1	,347[*]	,172	-,397[**]	,081	,194	-,065	-,036

Morphological correlations

	the trees	men	I beat	we beat	twenty	sleeping	twenty one	they saw	to go	you saw	you beat
he has been seen	-,196	-,193	-,048	,046	-,295	,318[*]	-,292	-,145	-,517[**]	-,140	,157
	,214	,222	,763	,774	,057	,040	,061	,359	,000	,378	,320

References

1. Asadpour, Hiwa (2007). *The Linguistic Atlas of Western Azerbaijan*. Mashhad: Ferdowsi University Press.

2. Asadpour, Hiwa (2011). *The Computer Developed Linguistic Atlas of Azarbaijan-e Qarbi. Science and Research branch,* Islamic Âzad University, Tehran, Unpublished MA thesis.

3. Bloomfield, L. (2000). *Language*, Translator Haghshenas, A. M., First Publication, Tehran; University Press Center.
4. Breton, Ronlad. J. L. (1991). *Geolinguistics: Language Dynamics and Ethnolinguistic Geography*. Translated and Expanded by Harold F. Schiffman. Ottawa and Paris: University of Ottawa Press.
5. Chambers, J. K. (1995). *Sociolinguistic Theory*. Oxford: Blackwell.
6. Chambers, J. K. and Trudgill, P. (1980). *"Dialectology"*. Cambridge; Cambridge University Press.
7. Chambers, J. K. and Trudgill, P. (1998). *"Dialectology"*. Cambridge, Cambridge University Press.
8. Chambers, J. K., Trudgill, P. & Natalie Schilling – Estes (eds.) (2002). *The Handbook of Language Variation and Change*. Oxford: Blackwell.
9. Coulmas, Florian (2006). *Sociolinguistics; The Study of Speaker's Choices*. Cambridge: Cambridge University Press.
10. Goebl, H. (1982). *Dialektometrie; Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie*, volume 157 of Philosophisch-Historische Klasse Denkschriften. Verlag der Österreichischen Akademie der Wissenschaften, Vienna. With assistance of W.-D. Rase and H. Pudlatz.
11. Goebl. H. (1984). *Dialektometrische Studien. Anhand italoromanischer, r¨atoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, volume 191, 192, 193 of *eihefte zur Zeitschrift f¨ur romanische Philologie*. Max Niemeyer Verlag, T¨ubingen. With assistance of S. Selberherr, W.-D. Rase and H. Pudlatz.
12. Goebl, H. (1993). *Probleme und Methoden der Dialektometrie: Geolinguistik in globaler Perspektive*. In Viereck, W. (ed.), Proceedings of the International Congress of Dialectologists, volume 1, p. 37-81, Stuttgart. Franz Steiner Verlag.
13. Gooskens, C. and Heeringa, W. (2004). *Perceptive Evaluation of Levenshtein Dialect Distance Measurements using Norwegian Dialect Data*. Language Variation and Change, 16(3):189–207.
14. Haugen, E. (1966). *'Dialect, Language, Nation'*. American Anthropologist 68:922-35. (1944) 'Standardization'. ELL:4340-2.
15. Heeringa, W. and Nerbonne, J. (2001). *Dialect Areas and Dialect Continua*. Language Variation and Change, 13:375–400.
16. Hudson, R. A. (1996). *"Sociolinguistics"*. Cambridge: Cambridge University Press.
17. Kerswill, Paul (2004). *Social dialectology/Sozialdialektologie*. In Klaus Mattheier, Ulrich Ammon & Peter Trudgill (eds.) *Sociolinguistics/Soziolinguistik. An international handbook of the science of language and societ*y, 2 nd edn., Vol 1. Berlin: De Gruyter. 22-33.
18. Jain, A. K. and Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, New Yersey.
19. Kessler B. (1995). *Computational Dialectology in Irish Gaelic*. In *Proceedings of the 7th Conferenceof the European Chapter of the Association for Computational Linguistics*, EACL, Dublin, pp. 60–67.
20. Labov, W. (1990). *The Intersection of Sex & Social Class in the Course of Linguistic Change*. Language Variation and Change 2: 205-54. Washington D. C: center for applied linguistics.
21. Labov, W. (1996). *The Social Stratification of English in New York cit., Washington D. C*: center for applied linguistics.
22. Lambert, W., R. Hodgson, R. Gardner and S. Fillenbaum. (1960). *Evaluational Reactions to Spoken Languages*. Journal of Abnormal and Social Psychology, 60, 44-51.
23. Margaret, Laing & Lass, Roger. (2006). *Early Middle English Dialectology: Problems and Prospects*. Unit d Kingdom of Great Britain and Northern Ireland.
24. Milroy, L. Milroy, J. (1992). *Social Network and Social Class: Toward an Integrated Sociolinguistic Model*. Language in Society 21: 1-26.
25. Petyt, K. M. (1980). *"The study of Dialect"*. London; The Trinity Press.
26. Orton, H, Sanderson, S. and Widdowson, J. (eds.) (1978). *"The Linguistic Atlas of England [474 maps]"*. Croom Helm.
27. Séguy, J. (1971), *"La relation entre la distance spatiale et la distance lexicale"*. In: Revue de Linguistique Romane, 35 (1971), p. 335-357.
28. Séguy, J. (1973). *Atlas linguistique de la France par régions, atlas linguistique de la Gascogne, complément du volume VI*. Paris, centre national de la recherche scientifique.
29. Sokal, R. R. and Rohlf, F. J. (1962). *The Comparison of Dendrograms by Objective Methods*. Taxon, 11:33–40.
30. Trudgill, P. (1980) *On dialect*. Oxford: Blackwell.
31. Trudgill, P. (1986) *Dialects in Contact*. Oxford: Blackwell.

3/22/2018