

## Quantitative sociolinguistics Methodology A Descriptive and Analytical Study

**Dr. Amin Mustafa Alshangiti**

*Assistant professor at the Institute of Arabic Teaching to Non-Arabic Speakers,  
Islamic University, Madinah, Saudi Arabia*

### Abstract

This research aims to present and review the *quantitative sociolinguistic* framework (Labovian Method) adopted in many sociolinguistic studies (including Arabic Language studies). In addition, it describes, in detail, the methods and means that commonly used to collect, organise and analyse sociolinguistic data. The research also emphasises the fact that sociolinguistic data mainly elicited by using a common method in sociolinguistics, i.e. tape recording. Two common settings related to tape recording method were chosen to be reviewed: personal interviews and group discussions. Moreover, for sampling sociolinguistic research participants, two methods of sampling were reviewed and discussed, i.e. *random sampling and judgment sampling*. This study argues that the latter method seems to be the only appropriate sampling method to use in the Arab world, due to the difficulty to approach Arab speakers without pre-arrangement. More importantly, the study reviewed two fundamental components of the quantitative sociolinguistic studies, which have to be correlated to each other, i.e. sociological and linguistic variables.

**Keywords:** Quantitative sociolinguistics, Labovian Methodology, sociological and linguistic variables.

## اللسانيات الاجتماعية الكمية : دراسة وصفية تحليلية

د. أمين مصطفى الشنقيطي

أستاذ مساعد بمعهد تعليم اللغة العربية لغير الناطقين بها

الجامعة الإسلامية - المدينة المنورة - المملكة العربية السعودية

### مستخلص البحث

يهدف هذا البحث إلى تقديم ومراجعة منهجية اللسانيات الاجتماعية الكمية (المنهجية اللابوفية)، والتي تبنتها العديد من الدراسات في اللسانيات الاجتماعية (والتي تشمل دراسات اللغة العربية في هذا المجال). بالإضافة إلى أن البحث تناول -بشيء من التفصيل- الطرائق المشتهرة المستعملة في جمع وتنظيم وتحليل بيانات اللسانيات الاجتماعية. وأكد هذا البحث على أن بيانات اللسانيات الاجتماعية تجمع بطرق متعددة، ولكن من أشهرها وأكثرها شيوعا التسجيلات الصوتية. ولذلك راجع هذا البحث الهيئات والطرق المتعددة المتعلقة بالتسجيل الصوتي، وأشهرها المقابلات الشخصية وجلسات النقاش الجماعي. وقد تناولت الدراسة أيضا بالبحث والمراجعة طرق تحديد واختيار المشاركين في بحث اللسانيات الاجتماعية، وتحديدًا طريقتا الاختيار العشوائي لعينات المشاركين والاختيار المُحَكَّم لها. وتوصلت الدراسة إلى أن الطريقة الأخيرة تبدو هي الطريقة الوحيدة المناسبة في المجتمعات العربية لأنه يصعب التواصل مع أفراد المجتمع العربي لغرض المشاركة في المقابلات الشخصية من غير ترتيبات تسبق إجراء تلك المقابلات. وهذه الدراسة أيضا راجعت مكونين رئيسيين في دراسات اللسانيات الاجتماعية الكمية، وهما المتغيرات الاجتماعية واللغوية.

**الكلمات المفتاحية:** اللسانيات الاجتماعية الكمية، المنهجية اللابوفية، المتغيرات الاجتماعية واللغوية.

## 1.0 Introduction

The main objective of this paper is to address in detail the *quantitative sociolinguistic* research methodology. The main focus of this study is to explicitly review the statistical method adopted by the American linguist William Labov, who is the leading figure of this methodology. This review study will discuss the method used to select the empirical sociolinguistic studies' informants. The study will also provide a review on how the informants' speech is sampled in this type of language empirical research. The study will review the independent (social) and dependent (linguistic) variables of the quantitative sociolinguistic framework.

## 2.0 Quantitative sociolinguistic method (Labovian Methodology)

In empirical research (whether in linguistics or any other subject), the validity and the importance of the information collected depends, primarily, on the methodology that the fieldworker uses to obtain that information. It is always challenging to choose and adopt a suitable and valid methodological framework for a study, especially when it involves collecting informants' dialectal speech (or the vernacular). Vaux & Cooper (2003: 178) identified three basic challenges associated with attempting to conduct fieldwork in dialectology: the first basic challenge facing the fieldworker is to identify his/her informants and maintain their help and cooperation. In addition, it is important that the informants feel comfortable speaking non-standard dialect, as the researcher can face difficulty in eliciting dialect data successfully, in face of the fact that most speakers feel that they have no non-standard linguistic features (ibid).

There are various sociolinguistic methods used to select samples and record their speech and choosing the appropriate method is, to a

large extent, dependent on the research aims, and objectives, that the fieldworker is trying to achieve (Milroy 1987: 28). It is worth mentioning that not all sampling methods are relevant to all speech communities. For instance, if we take social class as a variable in two different geographical areas, such as in Western speech communities, which have been the subject of extensive studies in language variation, and in Arabic speech communities, we will discover that this variable is mostly defined in terms of socioeconomic standards (e.g. income, occupation, etc.) in Western speech communities (cf. Milroy (ibid: 29). This approach towards social class is very common among sociolinguists, including Labov (1966), Wolfram (1969), Fasold (1972), Trudgill (1974), and Rickford (1986: 215). Trudgill (1974: 32) states that “social classes are not organised or sharply demarcated social groups, but rather aggregates of people with similar economic characteristics”. On the other hand, in many Arabic speech communities (especially non-urbanised ones, i.e. rural and Bedouin), this social class might be more usefully defined by non- socioeconomic factors, such as level of education, ethnicity, tribal affiliation etc. Therefore, it is very problematic to say that the correlation between linguistic variables and certain social variables should be applicable and typical for all speech communities, regardless of any differences between them (Al-Shehri 1993: 37f).

It is a fact, that the methodological framework adopted by William Labov, who was

“the leading figure in this field and pioneered work of this type, notably in his 1966 publication”<sup>1</sup> (Trudgill 2003: 71),

---

<sup>1</sup> *The Social Stratification of English in New York City.*

received more attention than any other study in the last century. The validity and importance of Labovian methodology, according to Trudgill (1998: 157), is that it proves that the language variation process is not a chaotic one<sup>2</sup>. In his study, Labov examined phonological variables, such as the rhoticity of the /r/ sound, and how the realisation of this variable, varied in the speech of the community under investigation. In his study, three social variables were examined: education, occupation and income. He identified four social stratifications, involved in the analysis and correlation between social and linguistic variables: lower class, working class, lower middle class, and upper middle class (cf. Labov 1966: 133ff). After the leading Labovian studies, many studies were conducted in a similar manner concerning different Western societies. For instance, Trudgill (1974) studied the social differentiation of English in Norwich'. This study examined the same social variables proposed by Labov (1966), in addition to three more variables: locality, housing scale (ownership, age, and type) and father's occupation. Then he proposed similar social stratifications to those previously proposed by Labov, with sub-divisions of those variables (cf. Trudgill 1974: 31ff).

Al-Shehri (1993: 39) argues that social class as a variable in language variation studies is more appropriately defined in socio-economic and education terms in the developed (highly industrialised) societies in the West. Moreover, the indicators proposed by Labov (1966) and Trudgill (1974), such as income, occupation and type of housing are very useful for identifying the social class scale in these societies, where economic changes in speech communities are clearly reflected in language variation. Therefore, the correlation

---

<sup>2</sup> See, for example, Labov (1966, 2001).

between social class, based on the above criteria and linguistic variables, is clear and easy to trace. In contrast, due to the difficulty of finding clear socioeconomic stratification, this correlation in the so-called 'Third World' societies (lowly industrialised) might be irrelevant in some cases, or not fruitful to examine in others. On the other hand, educational attainment and religious affiliation, for instance, might be more effective markers of social-class differentiations in Arab world speech communities.

## 2.1 Sampling the informants

### 2.2.1 Methods for sampling informants

In sociolinguistic studies, the informant sampling method is no less important than the information the fieldworker is intending to collect from his/her informants. Therefore, selecting an inappropriate method may have a negative effect on the reliability and validity of the study. Therefore, adopting a sampling method, that is suitable for the nature of the study, is an important factor to bring out the relation between research design and research objectives (Milroy 1987: 18). In sociolinguistics, there are two main widely-known sampling methods, *random sampling* and *judgment sampling*. Each of them has its own sociolinguistic objectives and adopting one of them should be based on what has been explained above.

The first method (random sampling) was first adopted by William Labov, in his ground-breaking study of English in New York: *The Social Stratification of English in New York City* (1966). The most remarkable aspect of his method, was that his sample frame, gave everyone in the speech community an equal chance to be selected for the study. This was aimed at resolving the representativeness problem (Trudgill 1984:

203). Labov's *sample frame* refers to any population list, which could include electoral registers and telephone directories. Milroy (1987:19) argues that William Labov in his innovatory work [1966] was by no means the first urban dialectologist to be sensitive to the need to give a representative account of urban speech, his sampling methods are, however, important and distinctive. This Labovian method is clearly held in high regard since it was developed until recent times: e.g. Chambers & Trudgill (1980); Hudson (1980); Trudgill (1984); Wardhaugh (1986); Milroy (1987). The best example of a study that adopted this method, is Peter Trudgill's study of English in his home city of Norwich, *The Social Differentiation of English in Norwich* (1974), although it was also adopted by a number of other studies.

Despite the fact that the Labovian sampling method gained high prominence, as it represents a very important proportion of language variation studies, it is not free from criticism associated with its implementation in the proposed speech community. For instance, this method is not without bias; for example, electoral registers do not include people under 18 years old, and telephone directories only include those people who have a subscription with the service provider. In other words, Labov had a role in the selection of his samples, and in the exclusion of those who did not fulfil his criteria (Milroy 1987: 19). This method has been abandoned by the majority of sociolinguistic studies in recent times, in favour of *judgment sampling*, as representativeness is less likely to be achieved with large populations with diverse members. In addition, *random sampling* presents difficulties in terms of constructing a well stratified and balanced sample; *judgment sampling* is therefore preferable in this respect. (see Milroy &

Gordon 2003: 24ff; Alessa 2008: 31).

Labov's sampling methods, which are relatively complicated, have been discussed and examined in terms of their suitability and validity. There is, also, a question concerning their validity in other disciplines outside linguistic studies (Trudgill 1984: 203). Moreover, Milroy (1987: 27) states that, ultimately, his method can, in actual fact, be described as *judgment sampling*, rather than *random sampling*, as although the Labov's sample size was large, he discarded the majority of his samples, because the sample members did not meet his criteria.

The *judgment sampling* method, on the other hand, seems more reliable when it is well-constructed, according to the researcher's judgments. The main principle of this method is that the researcher chooses the different types of informants he/she intends to study, and then looks for a quota of informants that fits his/her proposed criteria. Ultimately, the judgment sample should be rational and well-motivated (Milroy 1987: 26). Moreover, this sampling method –has become the standard operating procedure not only in dialectology but also in sociolinguistics (Bailey & Dyer 1992: 3). In other words, the *judgment sampling* method is more appropriate to those social groups that are well-defined and specifiable. In contrast, the *random sampling* method rarely produces valuable outcomes, in studies of this kind of social group (Milroy 1987: 27).

### 2.2.2 The researcher and the speech community

A good relationship between the researcher and the speech community, whose speech he or she intends to investigate, is extremely important, especially in the case of closed societies, such as Arab societies. Therefore, it plays a vital role in the fieldworker gaining access to these

community members, thus allowing him or her to interview or record the participants without experiencing doubt or mistrust. Milroy (1980: 80) emphasises the link between the researcher having good relations with the community under investigation, and the success of the fieldwork: "...the closer the fieldworker is matched to subjects, the more successful he or she is likely to be". In other words, success is less likely when a fieldworker, from outside the speech community, collects the data. Labov (1972b: 215) maintains that "the study of language in its social context can only be done when the language is 'known' in the sense that the investigator can understand rapid conversation".

It worth mentioning that working with assistants or 'insiders' is important when conducting research in a speech community that has different social classes. In his study of black English, *Language in the Inner City: Studies in the Black English Vernacular*, Labov (1972a) relied on two black researchers (Robins and Lewis) as fieldworker assistants or 'insiders', to conduct fieldwork with black informants in Harlem in New York. Nonetheless, working with fieldwork assistants, who linguistically and ethnically belong to the society is important, especially when the research concerns both genders and different ethnic groups. Therefore, if a fieldworker is looking to conduct fieldwork in any Arab community, for example, he or she should cooperate with an assistant of the opposite gender.

Being a male fieldworker in some Arab communities, such as Saudi Arabian communities is always problematic as, in many cases, the fieldworker is not able to fill his female quota. For instance, Al-Shehri (1993) in his study of Jeddah, states that the female quota

was underrepresented in his sample even though he used a female assistant. In other studies, such as Al-Jehani's (1985) study of Mecca and Khtani's (1992) study of Abha (in Asir Province), females were not represented at all. On the other hand, the task of a female fieldworker might be relatively easier, because they can easily access female informants, and should find it much easier to access male informants, than would be the case if a male fieldworker required access to female informants. For instance, Alessa (2008) in her study of Jeddah was able to easily access her female informants, and was, to a great extent, successful in accessing male informants, in addition to being helped by a male assistant. Her situation as a female resulted in "a fair representation of both sexes: 27 males and 39 females" (Alessa 2008:55).

In order the sociolinguistic researcher to achieve representativeness in his/her data collection, the 'social network'<sup>3</sup> concept is beneficial to employ, as developed by Milroy (1980)<sup>4</sup> using the 'friend-of-a-friend' approach.<sup>5</sup> This technique is based on broadening the network contacts. For instance, when the first-order network contact (my friend, for example) introduces me to another person (a second-order network contact), then the second one may refer me to a third one (a third-order network contact), and so on. This technique was useful in facilitating the finding of suitable participants, that met specific social criteria that the researcher was not able to access from his first-order network contact.

<sup>3</sup> The concept of the 'social network' as an analytic concept was firstly introduced by Barnes (1954) (Milroy 1980: 46).

<sup>4</sup> See also Russel (1982); Bortoni-Ricardo (1985); Jabeur (1987).

<sup>5</sup> This term was first introduced by Boissevain (1974).

## 2.2 Sampling the informants' speech

Many studies that have been carried out in the sociolinguistic field have paid special attention to vernacular speech: the level of speech that is produced spontaneously by speech community speakers. The most important feature of this kind of speech is that it represents the indigenous language of a speech community, which has the most important value of the natural speech of the speech community. Moreover, this kind of speech is considered to be in contrast with less natural speech varieties, such as the *standard* and the *lingua franca*. (Crystal 2008: 511).

Although there are different methods of collecting sociolinguistic data, the face-to-face social interview technique, is still the most common and effective method for eliciting sociolinguistic data (cf. Milroy & Gordon 2003: 57). This fact does not remove the common problem associated with this method, as mentioned above. The level of negative impact of this method on speech spontaneity might vary from one speech community to another, so this problem has motivated sociolinguists to design their interviews in a way that reduces the negative impact of this method. It is difficult to achieve speech spontaneity when collecting data by this method, as subjects often produce unnatural speech, or shift to a standard form, when they realise that they are being observed and tape-recorded by others.

William Labov coined the term 'observer's paradox' to describe the common major problem associated with eliciting the vernacular in a speech community. He explains this term, by stating that: "the aim of linguistic research in the community must be to find out how people

talk when they are not being systematically observed; yet we can only obtain these data by systematic observation” (Labov 1972b: 209). In order to overcome this problem, or at least to reduce its negative impact, different methods have been implemented by sociolinguists and fieldworkers, such as the anthropological technique of ‘participant observation’. This technique is based on the fieldworker participating with the group under investigation, and becoming a member of this group for a period of time. Thus, the fieldworker will become an ‘insider’ observer, not an ‘outsider’ one. This new status of the fieldworker will facilitate in minimising the attention of the informants on their speech (Trudgill 2003: 101).

The ‘pre-interview question’ is a well-known technique adopted by William Labov (cf. Labov 1966) and others, in order to obtain spontaneous speech during interviews. With the ‘pre-interview question’, the fieldworker aims to trigger the subject’s participation in an informal way by asking him/her about something he/she is willing and enthusiastic to talk about. Labov (1966) adopted the ‘danger of death question’,<sup>6</sup> while Trudgill (1974) asked his informants about something humorous.<sup>7</sup> The choice between these two ‘pre-interview questions’, seems to have depended on what interested the communities in New York and Norwich, at the time of the data collection.

In his study of Norwich English, Trudgill (1974) tried another technique to elicit vernacular spontaneity. This technique was based on

<sup>6</sup> “Have you ever been in a situation where you thought you were in serious danger of being killed... where you thought to yourself, “this is it?” (Labov 1966: 71).

<sup>7</sup> “Have you ever been in a situation, recently or sometime ago, where you had a good laugh, or something funny or humorous happened to you, or you saw it happen to someone else?” (Trudgill 1974: 51).

the ‘pre-interview conversation’ and entailed encouraging the informant to speak outside the context of the formal interview, or interacting with the informant while they were speaking to a third person (Trudgill 1974: 51). The importance of this technique might be generalised as applying to the vast majority of Arab speech communities, if not all of them, since these communities are highly sensitive to any kind of interview, especially with strangers. Moreover, Blom & Gumperz (1972) in their study of code-switching, implemented another method in order to avoid the side effect of face-to-face sociolinguistic interviews, and to elicit spontaneous speech from their informants. This technique or method aimed to record ‘spontaneous group conversation’ instead of recording individual informants, which is more formal. It should be noted that both methods (sociolinguistic interviews and group conversations) were adopted in the present study, and both are explained in detail below.

### **2.2.3 interview Sociolinguistic**

Labov (1984: 29f) argues that this method is the only systematic and effective way to elicit the valuable casual, speech that quantitative analysis demands. Although Labov’s statement is to a large extent true, there are structural limitations in the data collected using this method. One of the most important limitations of this method, is that the elicitation of some variants is very difficult, or sometimes impossible, to achieve, due to the existence of vernacular forms, that can only be elicited in specific social situations.

These particular variants are unlikely to be elicited through formal interviews; instead, they occur in specific social situations, such as when peers are speaking to each other. This particular limitation of

interviews can be identified at all levels of linguistic analysis (phonological, morphological, syntactic and discourse). Milroy (1987: 51ff) clearly addresses this in her study of *Inner City Belfast*. She states that eliciting the vowel sound for ‘meet’ and ‘meat’ was problematic, since some variations of this vowel occur only in spontaneous speech, and not in informal interviews. Moreover, the limitation of the analysis of the data elicited, goes further in some studies, when comparing the data elicited by interviews, to that elicited from unobserved spontaneous conversation; there is a debate about the reliability of the data elicited by the first technique, as compared to the latter. It has been claimed that the approximation to the vernacular of the data elicited by the interview method is relatively poor (Al-Shehri 1993: 51).

One of the most useful approaches which overcomes, or at least reduces, the limitations of the sociolinguistic interview method, is to combine it with another supplementary method, namely, ‘spontaneous group conversation’ as mentioned above (highlighted in more detail below). Despite the possibility of the above limitations of the sociolinguistic interview method, the amount, and the quality of, the data that this method produces, by tape-recording, means that it is still the most important method for eliciting accurate data, especially in terms of phonetic variation (Labov *et al.* 1972). It is, also, the most obvious and structured method for collecting sociolinguistic data, as it allows the fieldworker to steer the interview back in the right direction, when he/she feels that it is digressing. The relative ease of controlling the interviews, enables it to be led in a way that facilitates obtaining the required pre-planned data.

As for the interviews structure and topics, Labov (1984) applied

the concept of ‘conversational interview modules’ (conversational networks), which refers to a “group of questions focusing on a particular topic” (Labov 1984: 33) as a very structured example of ‘interview modules’. The most important feature of his conversational modules, is that they successfully engaged with the informants, as a result of choosing topics that addressed the previous experience of his informants. Moreover, the questions were designed to shift from one module to another in a systematic manner. The ‘conversational interview modules’ technique is very useful, because it allows the fieldworker to establish the interview with a good engagement with his/her informants, and then move on systematically and ‘smoothly’, from one module to the next. However, Labov’s ‘network modules’ are not necessarily appropriate for all speech communities. Therefore, the fieldworker should design his/her interviews in a way that suits the informants. It should be emphasised, that some of Labov’s subject modules, such as the girls fighting and dating modules, are inappropriate to most, if not all, Arab communities.

### **2.2.2 Group discussion**

One of the most important characteristics of this method, is that it, usually, provides a high level of spontaneous speech as a result of its essence, whereby two or more people gather to discuss particular issues. The collective and reactive nature of this method, is expected to reduce (to a minimum) the speech–recording formality, which is one of the main problems of the interview method. Moreover, it is anticipated that the interactions between the parties involved in the discussion (including the fieldworker) will distract attention from the main role of the fieldworker as an observer of the speech behaviour, and will distract

informants from the fact that they are being tape-recorded. Furthermore, this method allows the fieldworker to notice the linguistic differences between the speech of an individual (in the individual interview) and when the individual interacts with a group of people (in the group discussion).

The fieldworker can have two roles in group discussion sessions. He/she may be an observer of a group of participants who carry out their conversation on a particular topic, and while he/she is watching, he records their speech and intervenes when necessary. This method has the advantage that the fieldworker has the chance to concentrate, and be more aware of the different linguistic behaviours that the members of the group are demonstrating. The disadvantage of this method is that when people realise there is someone (the fieldworker) sitting and observing them, and recording their speech, they will be, to some extent, subject to a sort of formality. Alternatively, in addition to observing and recording spontaneous conversations in group discussion sessions, the fieldworker can be involved in the discussion as one of the group. This technique is known as 'participant observation', which is referred to as "a process in which the observer's ' presence in a social situation is maintained for the purpose of scientific investigation. The observer is in a face-to-face relationship with the observed, and, by participating with them in their natural life setting, he gathers data" (Schwartz and Schwartz 1955: 344).

This method has a very important advantage, which is that formality will be reduced to an absolute minimum by the observer being one of the group. Therefore, eliciting vernacular speech with a very

limited likelihood of a shift to formal speech, is a great benefit when compared with the disadvantages of this method. The disadvantages include the possibility of the fieldworker concentrating less on linguistic elements due to his/her emotional involvement with his/her participants; such emotional involvement could detract his/her attention from observing linguistic elements, and from interacting with other members in the conversation. On the other hand, the deep involvement and empathetic relationship of the observer with the subjects, helps him/her to understand their life and social behaviours more deeply, which adds very important validity and meaningfulness to his/her data (ibid: 350).

### **2.3 Social variables**

In this research, four sociological variables are defined and reviewed below, i.e. age, educational attainment, ethnicity, and gender. The main factor behind choosing these particular variables, was the assumption that they would be suitable for all Arabic-speaking communities.

#### **2.3.1 Age**

Studying ‘age’ as a sociolinguistic variable, in order to correlate different age groups with linguistic variables in a speech community, seems to have been one of the most frequent social variables studied in this field, since Labov’s ‘inspirational’ study of the speech community in New York City (Labov 1966). Although this variable has been extensively examined in the field, it –by itself has no explanatory value; it is only when examined in the context of its social significance as something reflecting differences in life experiences that it becomes a useful analytical construct (Milroy & Gordon 2003: 39). The importance of studying age in sociolinguistic studies, is not only due to its correlation with the linguistic variation in a certain language or dialect,

but this social variable also plays an important role in one's mastering of a dialect in the case of shifting from one dialect to another, according to Chambers (1995: 85). He claims, that once people are over 14 years of age, it is difficult for them to acquire a new dialect, while the best age for acquiring a new dialect is under seven years of age, as children of this age are able to acquire native-like proficiency in the acquired dialect.

There are different approaches in the variationist literature, regarding classifying age groups, in order to investigate linguistic variation between different age groups. One of these approaches, involves considering chronological age as a grouping 'instrument'. The other approach suggested by Eckert (1996), is to group speakers according to their life stages: childhood, adolescence, and adulthood. Eckert (ibid: 156) states that the life stages approach is more appropriate than the chronological one, [as]other aspects of the passage through life are less specifically tied to chronological age and more tied to life events, such as changes in religious status (bar and bat mitzvah, baptism), institutional status (first day of school, retirement), family status (marriage, first child), legal status (naturalization, first arrest), and physiological status (loss of the first tooth, onset of menses). These events in turn are associated with life stages: childhood, adolescence, young adulthood, middle age, old age. It is these general life stages that are most frequently invoked to explain behaviour.

William Labov introduced two constructs (Trudgill 2003: 9) for analysing age-related linguistic change: apparent-time and real-time (Labov 1966). The first term, 'apparent-time', refers to studying language

variation and change in a specific speech community, by comparing the speech of older speakers with younger ones. It is based on the assumption that in the dialect of the community, change is manifested in the speech of different generations, as older speakers use old forms and younger speakers use newer ones (Trudgill 2003: 91). In other words, this method aims to study the “distribution of linguistic variables across age levels” (Labov 1994: 45f). The main objective of the other term, ‘real-time’, is to examine language variation and change at a particular point in time, in a specific speech community, with the fieldworker returning years later to do the same study on the same speech community. The aim is to identify the changes that have occurred in the speech community in the period of time since the initial fieldwork was conducted (Trudgill 2003: 109).

The main problem associated with the ‘apparent-time’ method is ‘age grading’. This speech behaviour occurs when speakers in a community change their speech behaviour as they get older, and yet these alterations are repeated in every generation. For example, some speakers in a speech community modify their linguistic behaviour towards the acrolect when they reach middle age, and then, gradually, reach the prestigious level of speech by retirement age (ibid: 6). Labov (1994: 73) suggests the second method (real-time technique) to overcome the possibility of age grading occurring. He argues that “the obvious answer to the problems involved in the interpretation of apparent time would be to rely upon observations in real time, that is, to observe a speech community at two discrete points in time”.

He identified two ways to elicit ‘real-time’ data (ibid). The first and easiest method is to compare the earlier speech community study

results, with the results of the current study. The second approach for obtaining ‘real-time’ data, involves reinvestigating the same speech community that was investigated years previously. The fieldworker should replicate the methods used in the earlier study as closely as possible, with the same informants or others. For example, Anders Steinsholt used this method when he conducted a dialect research study on the Norwegian community of Hedrum in the 1930s, before returning to the same speech community to do a similar one in the 1960s (Trudgill 2003: 109). Trudgill (1988) did the same when he revisited and studied the Norwich speech community, after conducting a study in 1974, which was based on the ‘apparent-time’ method. It seems that the ‘apparent-time’ method is more practical than the ‘real-time’ method, as the latter requires years or decades to allow the researcher to achieve his final findings, while the results of the former are available quickly after conducting and analysing the data. Furthermore, the results of the ‘apparent-time’ method data may be compared with ‘real-time’ data (Al-Shehri 1993: 61).

### 2.3.2 Gender

Regardless of the differences between the two terms concerning males and females, i.e. gender and sex, as the first is associated with social status, while the other is associated with biological context, male and female linguistic variation has been extensively highlighted by almost all sociolinguistic studies. This necessarily indicates the importance of studying gender-related linguistic variation in any speech community that has special linguistic properties. Labov (2001: 263) demonstrates explicitly that gender comes in different forms, and has a profound impact as a social variable in any speech community. The influence of

gender indicated by Labov may result in language variation at different levels; this has been addressed by many studies, including Trudgill (1972), Cameron & Coates (1985), and Eckert (1989) to mention only a few.

It seems that the studies, especially Western studies that dealt with gender-related linguistic variation, have concentrated on standard and prestigious versus non-standard or vernacular speech between males and females. Moreover, the stable linguistic variants usually show clear gender-related differentiation, when the production of these variables is analysed statistically. For instance, in English, the variable ‘-ing’ is a good example, where many studies have examined the gender-related differentiation in the production of this variable. These studies were conducted in different English-speaking communities and came to the general conclusion, that female speakers have a greater tendency than males to use the standard variant (ɪŋ) rather than the non-standard variant (ɪn) (see, for example, Fischer 1958; Labov 1966; Wolfarm 1969; Trudgill 1974). Furthermore, for the English interdental fricatives (θ) and (ð), women avoid using the non-standard variants (t) and (d) in some areas according to different studies, e.g. Labov (1966) in his study of New York and Anshen (1974) in his study of North Carolina.

The various degrees of linguistic variation between males and females are due to “the combination of economic, social and to some extent physical segregation by sex” (Francis 1983: 44). According to Milroy (1980: 112), it is a very common finding in urbanised Western speech communities that women are “approximating closer to the prestige pattern and style-shifting more extensively than men”. As a result of this general finding, Labov considered women to be the

initiators of linguistic change in a speech community, if not by themselves, by their direct influence on their children during the early age of language acquisition when children are forming linguistic rules (Labov 1972b: 302f).

Arabic studies, however, which have examined gender as a sociolinguistic variable have come to the opposite conclusion. In other words, men's speech is closer to standard variants than that of women in Arab speech communities. For instance, men approximate closer to the standard variant of ق (q) than women, as reported by Sallam (1980) and Schmidt (1986) in their studies of Egyptian Arabic spoken in Cairo. In Amman, men have a greater tendency to use prestige variants than women, according to Abdel-Jawad (1981). A similar finding has been demonstrated by different studies on different Arabic speech communities, e.g. Bakir (1986: Iraqi Arabic spoken in Basrah); Kojak (1983: Syrian Arabic); Wahba (1996: Egyptian Arabic spoken in Alexandria). The finding of the previous studies that contradicts the general finding of Western studies, mentioned above seems to be due to the diglossic situation of Arabic-speaking communities. In other words, it conforms with the local varieties (dialects) being considered as a low variety, while Classical/Modern Standard Arabic is seen as the prestigious (high) variety (Alessa 2008: 50).

Ibrahim (1986), supported by others, including Abdel-Jawad (1987) and Bakir (1986) proposes a new categorisation in this regard. He demonstrates that in Arabic-speaking communities, there are prestigious local varieties (supra-dialectal low), which are autonomous from Standard/Classical Arabic (Ibrahim 1986: 120). This pattern might lead to the general finding of Western studies that women's tendency

to approximate to prestigious norms is, generally, higher than that of men. This analysis to some extent conforms with Bakir's (1986) study of Basrah Arabic and Abu-Haidar's (1989) study of Baghdadi Arabic. Both studies came up with a similar finding, that women in both speech communities approximated to the prestigious variety, regardless of the direction of the approximation, which is in the direction of colloquial Iraqi in the first study and in the direction of Standard Arabic in the latter (Alessa 2008: 50-51). Chambers (1995: 144f) generalises the tendency of women, whether in the West or in the East, to approximate to standard varieties, and argues that:

When the linguistic situation in the Middle East is re-analysed in this way, taking into account the social ramifications of diglossia, the discrepancy between male and female responses in Middle Eastern and Western societies disappears, although the socio-cultural organization differs remarkably from the Western world, the sociolinguistic behaviour is essentially the same; women use more standard forms than men in the same social group in both worlds. The female advantage in verbal abilities apparently overrides the socio-cultural differences.

There is another view, which might be considered as a third approach towards the impact of gender on language variation in Arabic-speaking communities. It is based on relating the language variation to outside factors rather than gender as the determiner of language variation. In a study on Tunisian Arabic spoken in Korba, Walters (1991: 219) ascribed the level of language used by both genders to the choices that make sense in the context of these speakers' lives, the varieties of language to which they have access, and the social options available to them. Moreover, in Jabeur's (1987) study on Tunisian Arabic, spoken

in Rades, he argues that the speech differentiation in his speech community is not ascribed to gender essentially, but to other factors, such as the interaction between male and female speakers, educational opportunities, and socio-cultural changes. He found that due to cultural change, young males and females interact face-to-face in many social situations, and therefore their speech approximation is similar, e.g. their similar approximation to the (aj) and (aw) variants.

It seems that taking into account outside factors, such as social, historical, cultural, and ideological factors, is very important when correlating gender as a social variable with different linguistic behaviours. Therefore, Jabeur's general conclusion, which is supported by studies on different Arabic-speaking communities, such as that of the Najdi community in Jeddah, Saudi Arabia (Alessa 2008) and the Fallahis speech community in Karak, Jordan (El Slaman 2003), might be applicable to all Arabic-speaking communities. Milroy & Gordon (2003: 108) state that: "gender affects language differently in different generations because of various life experiences and gendered language differences index salient intra-community social categories which need to be uncovered by researchers rather than treated as previously given".

### 2.3.3 Education

The significance of studying the effect of the level of educational attainment, as a manifestation of verbal contact, on language variation, is due to the considerable importance of studying 'contact' itself as an important factor of language change. This factor (contact) has been intensively highlighted by numerous sociolinguistic studies. Jespersen (1946, cited in Chambers 1995: 242) states that "the most important

cause of language splitting into dialects is not purely physical, but want of communication for whatever reason". Labov (2001: 805) emphasises the importance of face-to-face interaction and argues that the lack of participation of African-Americans in the sound changes in his speech community is due to the "decreasing frequency of face-to-face interaction with speakers of the mainstream local dialect".

In the Arab world, where the percentage of illiteracy is very high<sup>8</sup>, the level of education is expected to be reflected in one's speech behaviour. Various sociolinguistic studies have investigated to what extent the level of education may have a direct impact on language variation. For instance, studies of different Jordanian Arabic speech communities, e.g. Abdel-Jawad (1981); Al-Khatib (1988); Kanakri (1988); El Salman (2003), link the use of the Standard Arabic sound (q) to the level of education that the speaker has attained. Al-Wer (1991: 52) emphasises the importance of the level of education of the speaker, which is an indicator of the amount of contact that occurs between him/her and the outside community.

#### 2.3.4 Ethnicity

There is no consensus on the definition of ethnicity and the elements that this term might include. Owens (2001: 434) studied this social variable in the Arab world, and argues that it refers to "any of a number of social parameters by which, non-national social groupings are distinguished, including religion, shared history, skin colour, kinship, lineage and place of origin. The relevant criterion or criteria defining ethnicity may differ from place to place".

<sup>8</sup> According to UNESCO (cited in Magin 2010), "40% of those over 15 years of age – nearly 70 million people are illiterate". See <http://www.gial.edu/documents/gialens/Vol4-2/Magin-Arab-Illiteracy.pdf>

Fishman (1977: 17) insists on paternity as an important element that constructs ethnicity; therefore, he narrowly defines it as being “in part, but at its core, experienced as an inherited constellation acquired from one’s parents as they acquired it from theirs, and so on back further and further, ad infinitum”. According to Bassiouney (2009: 98), Owen’s definition is broader than Fishman’s, including religion in the definition of ethnicity is problematic. She argues (ibid), that including religion when studying ethnicity in the Arab world might be politically charged, as this may not “reflect the way that people perceive themselves”. She exemplifies her view with the situation in Egypt, where people tend to perceive themselves as Egyptians (who have ancient history), rather than perceiving themselves as Copts or Muslims.

Owen’s statement that the criterion/criteria defining ethnicity may differ, from place to place, seems to be very true in the Arab world, where the elements that define ethnicity may differ from those in the West, where the culture and religion are different. For instance, religious affiliation (Sunni or Shiite) is a core criterion in defining ethnicity in a country like Iraq, especially since the Shiites took power after the collapse of Saddam’s regime, following the American-led invasion of the country. This criterion is irrelevant in other Arab countries, such as Mauritania, where the population of its original inhabitants are almost 100% Sunni Muslims.

Hall-Lew (2010: 458) argues that the categorisation of the term ‘ethnicity’ and its related term ‘race’ is constructed in a similar way to any other social category, e.g. gender and class, in many studies such as Fishman (1989), Fought (2006), Eckert (2008), and Becker & Coggshall (2009). Moreover, the term ‘ethnicity’ is associated with

shared aspects of a specific group of people, e.g. culture, religion, and heritage. The term ‘race’, on the other hand, is problematic, according to Hall–Lew (ibid), as it is “constructed with greater reference to perceived physical similarities, such as skin colour or facial features, which can vary widely within ethnic groups”.

Studying ethnicity as a social variable correlating to language variation and change is important in the world as a whole, and especially in the Arabic-speaking communities that are ethnically diverse. Bassiouney (2009: 99) emphasises the importance of studying ethnicity in multi-ethnic Arab communities, stating that “in the past century the Arab world has been in a state of flux for different reasons, some of them political and some economic. We definitely need more studies that examine variation between different ethnic communities in the Arab world”. It is worth mentioning here that although there are a number of multi-ethnic Arab communities, only a few have attracted the attention of researchers. One of these multi-ethnic communities that have been linguistically studied exhaustively is Jordan. The demographic situation in Jordan is very interesting, with two large nationalities (Jordanians and Palestinians) living together in a small country. Although both communities share the same religion, and the Arabic varieties spoken by the two are very similar, the Jordanians and Palestinians conceive themselves as being of different ethnicities (ibid).

#### **2.4 Linguistic variables**

The linguistic variable, as a sociolinguistic term, is sometimes known as a sociolinguistic variable, was initially developed by William Labov in his early work on variation theory and secular linguistics (Trudgill 2003: 82). Since then, correlating the ‘linguistic variable’ with different social

variables, e.g. gender, age, class, etc., has become a main part of linguistic analysis in the sociolinguistic field. Fasold (1990: 224) defines the sociolinguistic variable as

“a set of alternative ways of saying the same thing, although the alternatives [variants] will have social significance”. In most cases, linguistic variables are phonological, while the occurrence of lexical and grammatical variables is relatively less frequent (ibid). In terms of the correlation between linguistic variables and social variables, two terms can be found in the field of sociolinguistics: dependent and independent variables.

The dependent variable literally means that the occurrence of this variable depends on another factor (the independent variable(s)). The dependent variables are the linguistic variables, because the occurrence of the latter is dependent on the independent variables, which are the social variables. Hatch & Lazarathon (1991: 63) point out that the dependent variables (linguistic variables) are those can be measured or quantified, while the independent variables (social variables) are those that the researcher, or the fieldworker, supposes may have an impact, or be related to dependent variables.

### 3.0 Conclusion

This study presented and reviewed the quantitative sociolinguistic framework adopted as the methodological framework by the American linguist William Labov. It described, in detail, the methods and means that are used to collect, organise and analyse the quantitative sociolinguistic data. This study also reviewed and described two settings that are commonly used in this type of linguistic studies: personal interviews and group discussions. In addition, two methods of

sampling were reviewed and discussed, i.e. *random sampling* and *judgment sampling*. It was argued, in this paper, that the latter method, seems to be the only appropriate sampling method to use in the Arab world, due to the difficulty, if not impossibility, to approach Arab speakers without pre-arrangement. This is clearly due to the lack of openness in Arab communities, and the unfamiliarity with this type of empirical research. Finally, this research reviewed four social variables: age, education, ethnicity and gender, that are meant to be corelated with the linguistic variables.

### References

- ABDEL-JAWAD, H. R. 1981. *Phonological and social variation in Arabic in Amman*. Ph.D.Thesis, University of Pennsylvania.
- ABDEL-JAWAD, H. R. 1987. Cross dialectal variation in Arabic: competing prestigious forms. *Language in Society*, 3, 359-368.
- ABU-HAIDAR, F. 1989. Are Iraqi Women More Prestige Conscious than Men?: Sex Differentiation in Baghdadi Arabic. *Language in Society*, 18, 471-481.
- ALESSA, A. M. 2008. *Najdi speakers in Hijaz: a sociolinguistic investigation of dialect contact in Jeddah*. Unpubl. PhD thesis, University of Essex.
- AL-JEHANI, N. M. 1985. *Sociostylistic stratification of Arabic in Makkah*. Unpub Ph.D., University of Michigan.
- AL-KHATIB, M. A. A. 1988. *Sociolinguistic change in an expanding urban context : A case study of Irbid City, Jordan*. Unpubl. PhD thesis, Durham University.
- AL-SHEHRI, A. S. A. 1993. *Urbanization and Linguistic variation and change: a sociolinguistic study of the impact of urbanization on the linguistic behaviour of urbanized rural immigrants in Hijaz, Saudi Arabia*. Unpubl. PhD thesis, University of Essex.
- ANSHEN, F. S. 1974. *Speech variation among Negroes in a small Southern community*. Unpubl. PhD thesis, New York University.
- BAILEY, G. & DYER, M. 1992. An Approach to Sampling in Dialectology. *American Speech*, 67, 3-20.
- BAKIR, M. 1986. Sex Differences in the Approximation to Standard Arabic: A Case Study. *Anthropological Linguistics*, 28, 3-9.
- BASSIOUNEY, R. 2009. *Arabic sociolinguistics: topics in diglossia, gender, identity and politics*, Washington, DC, Georgetown University Press.
- BECKER, K. & COGGSHALL, E. L. 2009. The Sociolinguistics of Ethnicity in New York City. *Language and Linguistics Compass*, 3, 751-766.

- BLOM, J. & GUMPERZ, J. J. 1972. Social meaning in linguistic structures: code-switching in Norway. *In: GUMPERZ, J. J. & HYMES, D. H. (eds.) Directions in sociolinguistics: the ethnography of communication.* New York: Holt, Rinehart and Winston.
- BOISSEVAIN, J. 1974. *Friends of friends: networks, manipulators and coalitions,* Oxford, Basil Blackwell.
- BORTONI-RICARDO, S. M. 1985. *The urbanization of rural dialect speakers: a sociolinguistic study in Brazil,* Cambridge, Cambridge University Press.
- CAMERON, D. & COATES, J. 1985. Some problems in the sociolinguistic explanation of sex differences. *Language & Communication, 5,* 143–151.
- CHAMBERS, J. K. 1995. *Sociolinguistic theory: linguistic variation and its social significance,* Oxford; Cambridge, Blackwell.
- CRYSTAL, D. 2008. *A dictionary of linguistics and phonetics,* Malden, MA, Blackwell Publishing.
- ECKERT, P. 1989. The whole woman: Sex and gender differences in variation. *Language Variation and Change, 1,* 245–267.
- ECKERT, P. 1996. Age as a Sociolinguistic Variable. *In: COULMAS, F. (ed.) The Handbook of Sociolinguistics.* Oxford; Cambridge: Blackwell
- ECKERT, P. 2008. Where do ethnolects stop? *International Journal of Bilingualism, 25–42.*
- EL SALMAN, M. A. M. 2003. *Phonological and morphological variation in the speech of Fallahis in Karak (Jordan).* Unpub Ph.D, Durham University.
- FASOLD, R. W. 1972. *Tense marking in Black English: a linguistic and social analysis,* Arlington, Virginia., Center for Applied Linguistics.
- FASOLD, R. W. 1990. *The Sociolinguistics of Language,* Oxford, Basil Blackwell.

- FISHMAN, J. A. 1977. Language and ethnicity. *In: GILES, H. (ed.) Language, ethnicity, and intergroup relations*. London: Academic Press.
- FISHMAN, J. A. 1989. *Language and ethnicity in minority sociolinguistic perspective*, Clevedon; Philadelphia, Multilingual Matters.
- FOUGHT, C. 2006. *Language and ethnicity*, Cambridge; New York, Cambridge University Press.
- FRANCIS, W. N. 1983. *Dialectology: an introduction*, London; New York, Longman.
- HALL-LEW, L. 2010. Ethnicity and Sociolinguistic Variation in San Francisco. *Language and Linguistics Compass*, 4, 458-472.
- HUDSON, R. A. 1980. *Sociolinguistics*, Cambridge; New York, Cambridge University Press.
- IBRAHIM, M. H. 1986. Standard and Prestige Language: A Problem in Arabic Sociolinguistics. *Anthropological Linguistics*, 28, 115-126.
- JABEUR, M. 1987. *A sociolinguistic study in Tunisia: Rades*. Unpubl. PhD thesis, University of Reading.
- KANAKRI, M. 1988. *Style and style shifting in the speech of educated speakers of Jordanian Arabic*. Unpubl. PhD thesis, University of Wisconsin-Madison.
- KHTANI, A. S. 1992. *The impact of social change on linguistic behaviour: phonological variation in spoken Arabic Asir, Saudi Arabia*. Unpubl. PhD thesis, University of Essex.
- KOJAK, W. 1983. *Language and Sex: A Case Study of a Group of Educated Syrian Speakers of Arabic*. Unpubl. MA diss., University of Lancaster.
- LABOV, W. 1966. *The social stratification of English in New York City*, Washington, D. C, Centre for Applied Linguistics.
- LABOV, W. 1972a. *Language in the inner city: studies in the Black English*

- vernacular*, Philadelphia, University of Pennsylvania Press.
- LABOV, W. 1972b. *Sociolinguistic patterns*, Philadelphia, University of Pennsylvania Press.
- LABOV, W. 1984. Field Methods of the Project on Linguistic Change and Variation. Sociolinguistic Working Paper Number 81. *In*: BAUGH, J. & SHERZER, J. (eds.) *Language in use*. Englewood Cliffs, New Jersey: Prentice Hall.
- LABOV, W. 1994. *Principles of linguistic change. Vol.1, Internal factors*, Oxford, Blackwell.
- LABOV, W. 1997. Resyllabification. *In*: HINSKENS, F., HOUT, R. V. & WETZELS, W. L. (eds.) *Variation, change and phonological theory*. Amsterdam; Philadelphia: Benjamins.
- LABOV, W. 2001. *Principles of linguistic change. Vol.2, Social factors*, Oxford, Blackwell.
- LABOV, W., YAEGER, M. & STEINER, R. C. 1972. *A quantitative study of sound change in progress*, Philadelphia, U.S. Regional Survey.
- MILROY, L. & GORDON, M. 2003. *Sociolinguistics: Method and Interpretation*, Malden, Blackwell.
- MILROY, L. 1980. *Language and social networks*, Oxford, B. Blackwell.
- MILROY, L. 1987. *Observing and analysing natural language: a critical account of sociolinguistic method*, Oxford, Blackwell.
- OWENS, J. 2001. Arabic Sociolinguistics. *Arabica*, 48, 419–469.
- RICKFORD, J. 1986. The need for new approaches to social class analysis in sociolinguistics. *Language & Communication*, 6, 215–221.
- SALLAM, A. M. 1980. Phonological variation in educated spoken Arabic: a study of the uvular and related plosive types. *Bulletin of the School of Oriental and African Studies*, 43, 77–100.
- SCHMIDT, R. W. 1986. Applied Sociolinguistics: The Case of Arabic as a Second Language. *Anthropological Linguistics*, 28, 55–72.

- SCHWARTZ, M. S. & SCHWARTZ, C. G. 1955. Problems in Participant Observation. *American Journal of Sociology*, 60, 343–353.
- TRUDGILL, P. 1974. *The social differentiation of English in Norwich*, Cambridge, Cambridge University Press.
- TRUDGILL, P. 1984. *Language in the British Isles*, Cambridge, Cambridge Univ. Press.
- TRUDGILL, P. 1988. Norwich revisited: recent linguistic changes in an English urban. *English World-Wide*, 9, 33–49.
- TRUDGILL, P. 1998. Language contact and inherent variability: the absence of hypercorrection in East Anglian present–tense verb forms. *In*: TRUDGILL, P. & CHESHIRE, J. (eds.) *The Sociolinguistic reader: Multilingualism and variation*. London: Hodder Arnold.
- TRUDGILL, P. 2003. *A glossary of sociolinguistics*, New York, Oxford University Press.
- VAUX, B. & COOPER, J. 2003. Linguistic Field Methods. <http://www.scribd.com/doc/19797207/2/Introduction>.
- WAHBA, K. M. 1996. Linguistic variation in Alexandrian Arabic: the feature of emphasis. *In*: ELGIBALI, A. (ed.) *Understanding Arabic: essays in contemporary Arabic linguistics in honor of El-Said Badawi*. Cairo, Egypt: American University in Cairo Press.
- WALTERS, K. 1991. Women, men and linguistic variation in the Arab world. *In*: COMRIE, B. & EID, M. (eds.) *Perspectives on Arabic linguistics III: papers from the Third Annual Symposium on Arabic Linguistics*. Amsterdam; Philadelphia: John Benjamins.
- WARDHAUGH, R. 1986. *An introduction to sociolinguistics*, New York, Blackwell.
- WOLFRAM, W. 1969. *A sociolinguistic description of Detroit Negro speech*, Washington, Center for Applied Linguistics.