

## Research Methods in Sociolinguistics (pp. 5-9)

(Juan Manuel Hernández-Campoy)

The development of Sociolinguistics has been qualitatively and quantitatively outstanding within Linguistic Science since its beginning in the 1950s, with a steady growth in both theoretical and methodological developments as well as in its interdisciplinary directions within the spectrum of language and society. Field methods in sociolinguistic studies have been motivated by the various research objectives pursued: sociological, sociolinguistic, or linguistic goals. The aim of this paper is twofold: (i) to provide a review of the theoretical movements within Sociolinguistics, and, on the basis of this review, (ii) to explore their consequences and implications on the research methods used in the field. This will be achieved by conducting both a retrospective synthesis of past developments and achievements, and an exploration of the current situation and of potential future developments.

### **1. The origins of sociolinguistics**

Since Currie (1952) used the term ‘sociolinguistics’ for the first time, the field has experienced an outstanding qualitative and quantitative growth within Linguistic Science. Currently, it is widely agreed that Sociolinguistics is an area of Linguistics concerned with the scientific study of the relationships between language and society, which entails practising a different way of doing linguistics that is very much influenced by work in the social sciences. It is empirical research — i.e. based on observation —, specifically focusing on how human beings actually use language in social interaction in real, everyday life situations and studies languages exclusively in their naturalistic social context (see Labov 1997; Trudgill 1983a).

Sociolinguistics is a multidisciplinary branch of linguistic knowledge that developed partly out of anthropology, partly out of ethnography, partly out of sociology, and partly out of dialectology as natural epistemological heritage, which conditioned both theoretically and methodologically this field and conferred it with an interdisciplinary stance. The work of Dell Hymes and his anthropology and folklore-based formation, John Gumperz and his interactional ethnographic studies, Joshua Fishman and his sociology of language, and William Labov and his linguistic variation, constitute the disciplinary matrixes

of this field. According to Figueroa (1994), the underpinnings of these three original strands in Sociolinguistics are philosophically-based: the work of Hymes is framed within relativism, Labov's variationist approach built on realism, and Gumperz' is portrayed in the context of intentionality and the interpretation of social meaning (see also Hymes 1972, 1974; Janicki 1990; Williams 1992).

The use of fieldwork methods in sociolinguistic research is a practice inherited from late 19th and early 20th century anthropological and ethnographically-oriented linguists — who studied Indian languages in USA (Franz Boas, Edward Sapir, or Leonard Bloomfield; see Murray 1998), native languages in Africa, Asia, or Australia (Alan Gardiner, Bronislaw Malinowski, or John Firth; see Robins 1971) — and European dialectologists, such as Georg Wenker, Jules Gilliéron, Karl Jaberg, or Jakob Jud (see, for instance, Chambers & Trudgill 1980; Davis 1983; Francis 1983). Their aim was to obtain linguistic data collected in the field, i.e. in natural environments of spoken language, just as people usually and casually meet and interact, rather than in an office (Canger 2001:779). This new orientation reflects the Labovian's distinction between traditional armchair linguistics — the theoretical and introspective study of language conducted in one's own office — and real-world linguistics — going out into the real world to collect data on language as used by ordinary people in their social context in everyday life. Labov claimed that analysts must not rely on their own intuitions since experience has shown that introspective judgements — even those made by informants themselves — may easily violate the reliability and validity of a linguistic generalisation when describing the state of a language (see Labov 1975a, 1975b, 1996). Theorisation without data, especially for taxonomic-type classification building, is conceived as sterile, since theory is solely understood to emerge from empirically collected data.

The methodological rigour of sociolinguistic research to guarantee empirical validity is sustained on the observer's paradox (see Labov 1972: 209, see also Davies 2001) and the principles of representativeness and generalisability (see Babbie 2001a; Bailey and Tillery 1999, 2004; Feagin 2002; D. Sankoff 2006; Trudgill & Hernández-Campoy 2007; Wolfram 2004) that apply to all social sciences research. A challenge in sociolinguistic research is the need to overcome the observer's paradox effect during the data collection process: the elicitation of authentic casual speech production with no contamination from the fieldworker's presence in order to avoid any alteration of the experiment results (Bell 2007; Cicourel 2006) and without violating ethical conditions (see Johnstone 2000a, 2001). Further,

as Bucholtz (2003) points out, the ethnographic method of participant-observation has become crucial in the fieldworker's endeavour to overcome the interference of observation on the activity being observed (see below).

No less important was the growing interest in Sociology, together with the influence of Marxist linguistics and the concern of educationalists and sociologists about poverty and social disadvantage as political issues in western industrialised countries. They tried to answer socially relevant questions, such as the relation between language and social class in Great Britain (the failure of non-standard-speaking children in schools), language and race in the United States (the failure in schools of black children for being users of Afro-American Vernacular English rather than of the standard), language and immigration in Western Europe, sex-ism in language, cross-cultural miscommunication (i.e., ethnic misunderstanding and racial disharmony), and the issue of language planning in multilingual post-colonial developing countries (see Bolton 1992; Dittmar 1976).

In addition to the epistemological heritage from social sciences (Ethnography, Anthropology, and Sociology) and Dialectology, Sociolinguistics was also a reaction against previous Chomskyan and Saussurean paradigms and traditional dialectological practices, which resulted in a new theoretical model, Variationism, as a prominent sociolinguistic sub-field influenced by the quantitative revolution. The origins of Variationist Sociolinguistics are to be understood as a challenge to the concept of the idiolect, the structuralist notions of *langue*/*parole* (language/ speech) and diachrony/synchrony postulated by Ferdinand de Saussure, the later generativist concepts of competence/performance proposed by Noam Chomsky, the Bloomfieldian notion of free variation, and the unrealistic theories and unreliable methodologies of the dialectological tradition. Linguists focused on micro-linguistics, the systematic homogeneity of *langue* and the speaker's competence, deliberately ignoring the macrolinguistic level with the orderly heterogeneous *parole* and the speaker's performance, and appealing to the Bloomfieldian notion of free variation as an explanation of any kind of linguistic variability (see Figure 1).

For Chomsky, the focus of study was the abstract system (competence), since linguistic performance was regarded as too disorderly and chaotic to be of any value in offering an understanding of language as a system (Baxter 2010: 118). In contrast, Labov regretted the exclusion of the study of actual sociolinguistic behaviour. He mentioned four distinct difficulties in the investigation of everyday

speech that made previous Saussurean and Chomskyan paradigms concentrate on langue and competence, and which condition the research methods to apply: (i) the ungrammaticality of everyday speech, (ii) variation in speech and in the speech community, (iii) difficulties of hearing and recording real speech, and (iv) the rarity of syntactic forms (Labov 1972: 183–259). In this way, it is not difficult to realise why in the past linguistic studies used to be of the armchair type.

The fact that most speech communities are to some extent socially and linguistically heterogeneous is a complexity which makes research much more difficult to any linguist wishing to describe a particular variety (Trudgill 1983a: 37).

Therefore, for many years during the end of the 19th and early 20th centuries linguists simply ignored this complexity. Dialectologists had focused on either the idiolect or the speech of rural informants, particularly that of elderly people with little education and little travelling experience, in small isolated villages, because they were very concerned with looking for the ‘real’ or ‘pure’ dialects. But, obviously, the idiolect — the speech of one person at one time in one style — was no more regular than the speech of the community as a whole, and ‘real’ or ‘pure’ homogeneous dialects turned out to be a fantasy. A monolithic linguistic system was unable to explain the fact that social structure could maintain any causal relation with the variability present in language. The obvious reaction against this theoretical model of language resulted in a shift from the fictional notion of systematically homogeneous to the orderly heterogeneous ‘speech community’.

After the Second World War, traditional dialectologists realised that confining dialect studies to rural areas meant ignoring the speech of the vast majority of the population, i.e., the speech found in large urban areas, which could not be investigated by applying the methods of traditional rural dialectology. Hence, starting the 1960s, some linguists became interested in macrolinguistics, under-taking the study of the complex reality of language in use in society considering the microlinguistic levels of analysis as linguistic variables. In this way, with the work of William Labov in the United States (see Murray 1998) and Peter Trudgill in England (see Hernández-Campoy 1993), Urban Dialectology appeared, conveying both a linguistic and a social function.

Finally, the study of the equational relationship between language and society through the correlation of extralinguistic factors (socio-demographic and/or con-text variables) with intralinguistic elements allowed Sociolinguistics to decipher the algorithm encrypting linguistic variation and social meaning and, consequently, to account for variability in language. The origins of Variationist Sociolinguistics did not ignore the neopositivist quantitative revolution. With their rigorous adoption of scientific methods, assuming determinism and the mechanistic nature of human behaviour, linguists' explicit positivist desire was to develop a quantified social dialectology where extralinguistic (mostly social) factors are capable, by themselves, of explaining entirely the establishment of laws, relationships and processes (see Bayley 2002; Paolillo 2001; Rietveld & van Hout 2006; D. Sankoff 2001). Labov's (1969) concept of variable rules and its mathematical implementation Varbrul (Cedergren & Sankoff 1974), for example, were developed to describe the predictable probability of patterns of variation (or choice) between alternative forms in language use, and the relationship between dependent (linguistic) and independent (extralinguistic) variables. These aims of explanation and prediction make sociolinguists focus not on what phenomena happen, where and how — which would solely be descriptive in intent — but on the reasons why they occur as they do. Likewise, in order to express accurately and plainly the results of their analysis, variationists demand from themselves the use of the language of mathematics and logic, validity and verifiability being the fundamental criteria and coincidence being conceived in terms of probability (see Hernández-Campoy & Almeida 2005: 10–11). Significance is here understood as the causal relationship between linguistic and extralinguistic variables in compliance of the principles of representativeness and generalisability (reliability and intersubjectivity) to pursue empirical validity. In this way, one can predictably guess the speech characterisation of speakers depending on their social background (class, age, gender, mobility, ethnicity, etc.), allowing sociolinguists to be a kind of omniscient observers in search for empirically valid sociolinguistic universals under the protection of the observer's paradox effect (i.e. Bayley & Lucas 2007; Hernández-Campoy & Almeida 2005; Paolillo 2001; Tagliamonte 2006).

[The Handbook of Linguistics \(p. 567; pp. 572-575\)](#)

[\(Mark Aronoff & Janie Rees-Miller\)](#)

#### **4. Micro-sociolinguistics**

Stated in very general terms, micro-sociolinguistics investigates how social structure influences the way people talk and how language varieties and patterns of use correlate with social attributes such as class, sex, age, and ethnicity. It thus strives to correlate dependent linguistic variables with independent social variables. Speech is socially emblematic in the sense that speakers by their choice of words, manner of pronunciation, and other stylistic features identify with others with whom they share social characteristics, such as socio-economic status, occupation, and education, but also place of residence, age, gender, and ethnicity. It is here that sociolinguistics relies most evidently on sociological theory. For, while it is intuitively clear that ways of speaking reveal social meaning, as immortalized in Bernard Shaw's famous play *Pygmalion*, it is much less obvious how individuals cluster into social groups, how social hierarchies are structured, and to what extent social hierarchies can be compared across speech communities, cultures, and polities.

#### **7. Macro-sociolinguistics**

“Speech community” is one of the basic notions of sociolinguistics, one which should not be used without an explicit definition. A common language distinguishes a social group in a sense, but the relationship between language and group affiliation is more complex than that. On one hand, where language is the only commonality, it does not really define a social group, and on the other, social groups may be defined in terms of language even though their members make use of more than one language. As Gumperz (1968: 220) was quick to point out, “regardless of the linguistic differences among them, the speech varieties employed within a speech community form a system because they are related to a shared set of social norms.” Linguistic rules are like social norms in that their validity is guaranteed by social aggregates. An important lesson of sociolinguistics is that linguistic rules in this sense extend beyond the traditional areas of structural patterning, that is, phonology, syntax, and the lexicon. Choices speakers make of speech varieties including different languages are subject to shared expectations and are meaningful. That behavior is meaningful is another way of saying that it is social, since it can be meaningful only by virtue of being governed by rules, and rules presuppose a social frame of reference.

The speech community has also been defined as an attitudinal community. Stereotypes and values accorded different speech forms are viewed as determining membership (Labov 1972b). A more traditional definition reminiscent of dialect studies is that of a socio-geographical community whose members are assumed to exhibit a high degree of homogeneity in speech (Halliday 1978). Other definitions are based on shared repertoires affiliated with a common mother tongue (Kloss 1966), and speakers' claim to membership (Coulmas 1996). Most generally, speech communities can be defined in terms of shared expectations and rules. Whether or to what extent a set of social rules is shared by individual speakers must be determined by surveys. Specific surveys designed to gather data on language characteristics as well as censuses that include questions about language use and proficiency are an important instrument of macro-sociolinguistic research, that is, the sociology of language. Results of such surveys are used, along with other data, to describe the linguistic composition of speech communities, their delimitation and interaction with each other. Such descriptions are a precondition for analyzing what societies do with their languages, and for recognizing the attitudes and attachments that account for the functional distribution of speech forms in a society.

If sociology is concerned with explaining social life, social relationships, socially meaningful behavior, and group affiliations, language must occupy the sociologist's attention, if only because language is one of the social possessions which most obviously reflects the internal differentiation of human societies. It lends itself easily to being used as a symbol of signaling commonality and marking boundaries. This is, perhaps, most obvious in multilingual speech communities, but there is plenty of linguistic diversity in monolingual communities too, and its investigation by sociolinguists has demonstrated time and again that it is just as socially meaningful. This kind of variation not only manifests subdivisions within communities and "community loyalty" (Milroy and Margrain 1980: 26), it is also characteristic of the individual speaker's language skills. Speakers can and do make choices. In their normal communication behavior they choose to pronounce words such as *mad* as one syllable [mæd] or two [me:ed]; they say "thanks" or "thank you" or they choose some other expression of gratitude; they choose formal and less formal styles, dialectal or standard varieties. The inability to do so marks a speaker as socially incompetent. On the basis of the extensive research into language variation that has accumulated since the late 1960s, it is safe to say that social differentiation requires linguistic

differentiation. If “requires” is understood here in a strict sense, it follows that explanations which can be derived from the relationship between language and society thus conceived will pertain, first and foremost, to language. “The social motivation of a sound change,” the title of an early article by Labov (1963), is indicative of this perspective. In this article, Labov investigated vowel centralization on Martha’s Vineyard, an island on the northeastern coast of the United States. He observed a generational change in the onset of the diphthongs in words such as white, wide, house, and how which he tried to correlate with social variables. In so doing he referred to social facts in order to explain linguistic phenomena, regarding social structure and stratification – in this case: occupational differences between farmers and fishers – as the cause of linguistic processes. However, attempts, such as Trudgill’s (1978), to limit the sociolinguistic enterprise to establishing causal links in this direction only have not been successful, because language and society are so indissolubly connected with each other.

At any given time, the linguistic differentiation of social groups is a fact that can be observed. It not only reflects social structure, but is also part of the social fabric and as such determines how individuals and groups interact with each other. Language provides the stage on which many social conflicts are played out (Haugen 1966, Edwards 1985, Nelde 1989). People are appreciated and discriminated against for the way they speak (Fishman 1989, Honey 1989); their attitudes towards their own languages and those of others are emotional (Williams 1974), sometimes as strongly as religious beliefs. They have ideas about linguistic goodness and purity (Jernudd, Shapiro 1989, Thomas 1991), what does and what does not belong to their language, how language should be taught (Stubbs 1976), etc. Language is frequently used as a defining characteristic of nationhood (Wardhaugh 1983) and ethnicity (Fishman 1989b), and it is associated with stereotypes about these (Le Page, Tabouret-Keller 1985). Languages are acquired and defended in order to achieve social goals (Cummins 1986). They have economic utility and are a factor of the transaction costs of economic processes (Coulmas 1992). Accordingly, they are perceived as the key to empowerment and success (Tollefson 1993). These beliefs and attitudes, no matter whether or not they are borne out by reality, are social facts in their own right which interact with other social facts, which in turn may have an effect on how languages evolve. Not only does language reflect social order, it contributes to its perpetuation and / or change. Deliberately guided language change has been promoted as a means to advance modernization, eradicate or reduce racial and sexual discrimination, and create

national cohesion (Cooper 1989, Weinstein 1983; see chapter 32). Policies designed to expand the geographical and / or social domains of a language – e.g., Swahili in post-colonial Tanzania – have created social facts (Calvet 1987, Pütz 1995). Dialects have broken away and been established as independent national languages for political reasons. For instance, Maltese, once a substandard Arabic trade jargon was accorded the status of Malta's national language (Hull 1994). Similarly, under German occupation during World War II, Luxembourgers reinforced the differences between Lëtzebuergesch and Standard German to turn their speech from a German dialect into a proper language (Kramer 1994).

Further, linguistic diversity, such as is characteristic of south Asia (Shapiro and Schiffman 1981) and many African societies (Herbert 1992), forces people to develop adaptive strategies such as creating and using a lingua franca (Calvet 1981). The multiplicity of languages is a social given for every member who is born into such a society. Here again language is a constituent of society rather than a mere reflection of its constitution. It interacts with other universals characteristic of human societies, i.e., technology (material culture), the organization of the social microcosm (kinship), the organization of the social macrocosm (politics) and the organization of the cosmos of beliefs (religion / ideology). The sociology of language explores these connections. Their relationships with micro-sociolinguistic issues are varied, some being more closely linked, others having a more distant connection. It has not proven feasible, however, to draw a sharp line between these two orientations.

Many questions can be investigated with equal justification within micro or macro-sociolinguistics. For instance, Uriel Weinreich's (1968) concern with language contact focussed on the traces that can be detected in linguistic systems of the contact and interaction of neighboring speech communities through their bilingual members. However, the preconditions and consequences of language contact involve a range of interesting phenomena, social and linguistic, which have both micro- and macro-aspects. "Contact linguistics" is now recognized as a branch of sociolinguistics (Nelde et al. 1995). The following can all be viewed as consequences of language contact: Language generation, i.e., pidginization and creolization (Mühlhäusler 1986, Bickerton 1992); language degeneration, i.e., language displacement (Dorian 1989); and novel patterns of language use, i.e., codeswitching (Myers-Scotton 1993a, 1993b). These and some other matters such as diglossia and bilingualism are discussed at length in chapter 21 and, therefore, need not be dealt with here. It is worth noting, however, that it is quite

impossible to say, without making arbitrary decisions, whether they should be treated properly in micro- or macro-sociolinguistics. The indissoluble connection between micro- and macro-issues has important repercussions for the question of a sociolinguistic theory.

## Language and Society (pp. 3-4)

(Raymond Hickey)

### **1.2 The development of sociolinguistics**

The development of sociolinguistics is bound up with the activity of American and British linguists since the early 1960s. First and foremost of these is William Labov who, in a pioneering investigation of the English of New York city published in 1966, arrived at many new conclusions concerning language variation and language change.

Labov stressed that 1) structural systems of the present and changes in languages of the past can be investigated in relation to each other, 2) language change can be observed in progress in present-day language varieties and 3) the fact that so-called 'free variation' was not in fact free at all but determined by deliberate, if not conscious, choices by speakers.

Labov further stressed the need to collect data reliably. The linguist must be aware that informants will show the following features in their speech: 1) style shifting (during an interview), 2) varying degree of attention, i.e. some speakers pay great attention to their own speech (so-called 'audiomonitoring'); in casual speech the attention paid is less, 3) degree of formality, determined by the nature of the interview, this can vary depending on the way informants react to the interviewer and the situations they are placed in.

The difficulty referred to above, namely that people's linguistic behaviour changes while being recorded, has been dubbed the observer's paradox by Labov. His answer to this problem was to develop the Rapid and Anonymous Interview in which informants were not aware they were being interviewed by a linguist. The essence of this technique can be seen by considering how Labov collected data on English in New York city. To begin with one should say that he was interested in the following linguistic variables: 1) the presence or absence of syllable-final /r/, 2) the pronunciation of the fricatives /2/ and /3/ and 3) the quality of various vowels. He chose two words in which these sound occurred, namely fourth floor, and then went around to a number of department stores in New York. Each of these was typical of a certain social class, and going on the assumption that employees use the pronunciation which holds for their typical customers, he could then examine the kind of English used in each store. To get samples without people knowing that they were acting as informants for a linguist, Labov checked in advance what items

were for sale on the fourth floor and then asked a store employee where he could find these items. After the individual responded 'on the fourth floor' he asked again, pretending that he did not hear the first time. This supplied him with a more careful pronunciation of the two words. Labov saw in this technique a means of gaining genuine pronunciations which were not spoiled by speakers' awareness of providing data for an investigating linguist. Of course, there are disadvantages to this method, above all the small quantity of data which can be gleaned at any one time and the inability to do a sound recording which one could listen to afterwards.