

## Collecting sociolinguistic data: Some typical and some not so typical approaches

Donna Starks & Zita McRobbie-Utasi

### Abstract

This article provides a survey of some data collection techniques employed in the analysing of social variation and language use. An outline of the methodology applied in early dialect studies and sociolinguistic interviews will be followed by a discussion of polling techniques including new and innovative rapid surveys used in data collecting. In addition, some of the findings obtained through employing different methods will be presented to illustrate their effectiveness in identifying and describing social variables.

This paper explores the three main types of data collection techniques: surveys, interviews, and polling techniques. An examination of survey methods used by traditional dialectologists will be followed by a review of interview techniques employed by researchers working within the variationist paradigm.<sup>1</sup> Particular emphasis is given to the sociolinguistic interview, the mainstay of modern sociolinguistic research. In the final section of the paper there is a survey of recent polling methods used to collect data on the social dimensions of language variation.

### Traditional dialectology

>From the beginning of the 19<sup>th</sup> century dialectologists have been working systematically on regional variation in language. Employing paper and pencil, and later tape recorders, researchers recorded differences in pronunciation, grammatical construction, and lexicon in the speech of rural inhabitants in France, Germany, England, Scotland, and America (Chambers & Trudgill, 1980, pp. 18-23; Petyt, 1980). Often, after decades of research, a monumental publication was produced which contained hundreds of dialect maps on which lines (called isoglosses) indicated the geographical limits of words, grammatical structures, and sounds. For example, although there are no current dialect maps of New Zealand or Australia, in the case of the former one could envisage such a map with isoglosses separating parts of Otago and Southland from the rest of New Zealand. These lines would indicate distinctive features of the area such as the presence of post-vocalic /r/ in words such as ‘*car*’, the vowel sound in words such as ‘*can’t*, *dance*’,

---

<sup>1</sup> The variationist paradigm is one of the main paradigms in sociolinguistics. It has been developed on the basis of recognizing that variability in language is systematic (Labov, 1966, 1972). Research within this paradigm focuses on

morphological features such as the past participle after ‘needs’ and ‘wants’ (eg, ‘*The cat wants fed*’), as well as a few lexical differences such as ‘*bach*’ in most of New Zealand vs ‘*crib*’ in Otago and Southland (Bartlett, 1992; Bayard & Bartlett, 1996, pp. 26-27; Bauer, 1997; Gordon & Deverson, 1998, pp. 126-129).<sup>2</sup> Although no other dialect areas have been identified within New Zealand, Bauer & Bauer’s (2000) current dialectology research on lexical variation may provide data for analysing additional regional variations in New Zealand English (see below).

Dialectologists employ two major techniques of data collecting, both of which involve 'direct probes' to elicit dialect forms (Wolfram & Schilling-Estes, 1998, p.21). Some of the earlier dialectologists used postal questionnaires mailed to selected individuals (eg, teachers); others used the 'on the spot phonetic transcription' method, by travelling from one rural community to another to collect information (see Chambers & Trudgill, 1980, pp. 18-23; Milroy, 1987, p. 10). The questionnaires were very detailed, often taking days, sometimes even weeks, to complete. For example, the questionnaire for the *Dictionary of American Regional English* contained over 1,800 questions (Wolfram & Schilling-Estes, 1998, p. 126).

There are several features characteristic of traditional dialectology. Earlier research tended to focus on a few older (mostly male) speakers (Chambers & Trudgill, 1980, p. 33) who lived all of their lives in the community in which they were born. The justification for this research attitude was the belief that these speakers had the ‘purest, most vernacular’ speech. However, because these speakers were living in rural areas, language use in urban centres was typically not considered. Another important feature of early dialectology was its theoretically biased approach. Dialectologists believed that through detailed documentation of the regional speech of older speakers living in isolated areas it was possible to show irregularity in language change (Kurath, 1972), thus refuting the then popular hypothesis concerning the regularity of sound change (the Neogrammarian Hypothesis).<sup>3</sup>

By the early 20<sup>th</sup> century, dialect research had shifted from a diachronic to a synchronic focus. It has become concerned with the distribution and variation of certain lexical or phonological items in population centres, rural and urban alike. While the social background of the various dialect

---

the relationship between variation in language and social factors, and on language change. The main paradigms in sociolinguistics are defined and described in Coupland & Jaworski (1997).

<sup>2</sup> There is considerable variability amongst Southlanders in the use of these dialect features. Many Southlanders use both the regional variant and the general New Zealand English form (Bartlett, 1992).

forms was also taken into consideration, this was only of secondary importance (Wolfram, 1997, p. 108). With increasing mobility and urbanisation, more complex dialectology methods have been devised. The *Linguistic Atlas of the United States*, for example, includes all population centres and represents individuals of different ages of three social types based on their social and educational backgrounds (Kurath, 1972).

### The sociolinguistic interview

A major shift in research techniques occurred with the publication of Labov's work on English in New York City (1966). His description of urban speech was based on a study of 88 individuals from a socially stratified random sample, consisting of male and female speakers from three age groups and four social classes (identified on the basis of education, occupation, and income). Labov showed that variation in the speech of the individual was a reflection of variation in the social group by illustrating how the most extreme case of stylistic variation in the use of /r/ by a single speaker was in conformity with the overall pattern exemplified in group scores of the different social classes (summarised in Chambers, 1995, pp. 18-21).

Labov's work on language use in New York City provided a blueprint for current methods of investigating variation in language use. As part of his research on the Lower-East side of New York City, he developed the sociolinguistic interview, the corner-stone of sociolinguistic research today. The sociolinguistic interview aims at eliciting linguistic data in different speech contexts. It comprises an informal part (consisting of free conversation) for eliciting vernacular or local use, and a formal part (consisting of a reading passage, word lists and minimal pairs<sup>4</sup>) to elicit various degrees of formal or standard language use.<sup>5</sup> Labov (1966) identified nine contextual styles from casual to formal, and associated all nine types with channel cues (ie, cues that signal change from one style to another). For example, by initiating a topic such as childhood games or traumatic life-threatening events the interviewer may achieve changes in the speech of the interviewee resulting in a less formal style, approximating or arriving at the desired more natural, vernacular speaking mode. The technique of inducing style change with this kind of prompt has been widely employed in sociolinguistic research (for example, Bayard, 1995;

---

<sup>3</sup> In Labov (1994) there is an in-depth discussion of the Neogrammarian Hypothesis.

<sup>4</sup>

Holmes & Bell, 1988, Appendix).<sup>6</sup> Thus the sociolinguistic interview usually starts with an informal free conversation, followed by increasingly formal language tasks that demand more attention to language use on the part of the respondent. The interviews often take up to two hours to complete (Holmes, Bell & Boyce, 1991).

Free conversation aims at eliciting 'natural speech', while the formal part of the interview is designed to elicit specific data that do not necessarily occur during the course of casual conversation. For example, as the vowel in the word '*fish*' occurs frequently enough in English speech, free conversation suffices should the linguist aim at examining this vowel.<sup>7</sup> However, certain other linguistic items may need careful interview design in order that sufficient data may be obtained for study. Word lists and minimal pairs can be constructed in such a way as to contain multiple tokens of the linguistic variable to be investigated.

In a sociolinguistic interview there is usually a minimum of five speakers per cell, with studies typically ranging from 48 to 120 respondents per community (Wolfram & Fasold, 1974).<sup>8</sup> Earlier studies relied on random sampling for data collection (Labov, 1966), whereas later researchers have tended to use judgement samples (see Wolfram & Fasold, 1974, p.38), or networking (Milroy 1987, pp. 35-36).<sup>9</sup> The sample is typically stratified on the basis of gender and age, and often includes social class (Labov, 1972; Trudgill, 1974) and/or ethnicity (Holmes, Bell & Boyce, 1991; Horvath, 1985). Of the social variables, age is often one of the most important. Bayard (1995, p. 65) claims that age appears to be the most important variable in New Zealand English, and Clarke (1991, p. 112) makes a similar statement about one variety of Newfoundland English.

An important feature to be acknowledged in connection with the sociolinguistic interview is the role of the interviewer. Some researchers have begun to explore the interviewer effect. For example, Trudgill (1986), has found that he was accommodating toward the speech of his

---

paradox" -- a basic methodological concern of sociolinguistic research -- denoting the dilemma as to how to obtain data on the way people speak when unobserved by the researcher who has to observe the speaker(s).

<sup>6</sup> For a detailed analysis of the sociolinguistic interview, we highly recommend Holmes & Bell (1988). This paper describes the interview schedule in the pilot sociolinguistic interviews for the Porirua study.

<sup>7</sup> Bell (1997, p.248), for example, was able to elicit a minimum of 50 tokens per speaker of the short (I) vowel in words such as '*fish*' in only 15 minutes of free conversation.

<sup>8</sup> Extraction of linguistic data from interviews is very time-consuming, making larger studies unfeasible. For a discussion of this, see Wardhaugh (1992, p.153).

<sup>9</sup> The units of social networks are 'pre-existing groups'. Researchers, instead of comparing groups of speakers, study relationships of individual speakers with other individuals (Milroy & Milroy, 1997, p. 59).

interviewees (ie, his own speech tended to replicate that of his subjects); however, few researchers have been able to assess the role of the interviewer effect. Bell & Johnson (1997)

schools in Christchurch have participated in their study. The results show that sound changes occur at different rates in different words. For example, although the change was complete for the word pair ‘







the pronunciation of one of the vowels in the speech of the interviewer, how the rapid survey can be used to analyse accommodation to different respondents during a brief encounter lasting less than a minute. For example, the findings showed that the interviewer tended to make slight adjustments according to the gender of the respondent: when interviewing a female respondent, the vowel tended to be more centralised than when interviewing a male respondent. The results illustrate the usefulness of polling techniques for analysing both the interviewer and the interviewee.

## Conclusion

It has long been recognised that in any language there exists a great degree of variation.<sup>13</sup> Members of a speech community are aware and make use of the different possibilities that the variability of language offers. In acknowledging the role of variability in language, linguists aim at identifying and accounting for a systematic pattern that governs the occurrence of the variants. This task can only be achieved by collecting representative data that reveal patterns of distribution with regard to the employing of the variants under investigation. The present study provides a brief survey of the most widely used techniques of sociolinguistic data collection. It discusses the approach of traditional dialectology, the use of the sociolinguistic interview, and the most frequently employed polling techniques including recent advances in rapid and anonymous surveys. The suitability of the various methods for studying variation in language are illustrated through their application in sociolinguistic studies.

## References

- Ash, S. (nd). Sampling strategy for the Telsur/Atlas project. Available: [http://www.ling.upenn/phono\\_atlas/sampling.html](http://www.ling.upenn/phono_atlas/sampling.html)
- Bailey, G., & Bernstein, C. (1989). Methodology for a phonological survey of Texas. *Journal of English Linguistics*, 22 (1), 6-16.
- Bartlett, C. (1992). Regional variation in New Zealand English: The case of Southland. *New Zealand English Newsletter*, 6, 5-15.

---

<sup>13</sup> The linguist Edward Sapir at the beginning of the 20<sup>th</sup> century stated that ...”everyone knows that language is a variable” (Sapir, 1921, p.147).

- Bauer, L. (1997). Attempting to trace Scottish influence on New Zealand English. In E. Schneider (Ed.), *New Englishes: Studies in honour of Manfred Görlach*. Amsterdam: Benjamins.
- Bauer, L., & Bauer, W. (2000). Nova Zelandia est ominis divisa in partes tres. *New Zealand English Journal*, 14, 7-17.
- Bayard, D. (1995). *Kiwitalk*. Palmerston North: Dunmore Press.
- Bayard, D., & Bartlett, C. (1996). 'You must be from Gorre': Attitudinal effects of Southland rhotic accents and speaker gender on NZE listeners and the question of NZE regional variation. *Te Reo*, 29, 25-45.
- Bell, A. (1997). The phonetics of 'fish and chips' in New Zealand: Marking national and ethnic identities. *English Worldwide*, 18, 243-270.
- Bell, A., & Johnson, G. (1997). Towards a sociolinguistics of style. *University of Pennsylvania Working Papers in Linguistics*, 4, 1-12.
- Boberg, C. (2000). Geo-linguistic diffusion and the U.S. Canada border. *Language Variation and Change*, 12, 1-24.
- Bourhis, R. Y. (1984). Cross-cultural communication in Montreal: Two field studies since Bill 101. *International Journal of the Sociology of Language*, 46, 3-47.
- Chambers, J. K. (1998). Inferring dialect from a postal questionnaire. *Journal of English Linguistic*, 26 (3), 226-246.
- Chambers, J. K. (1995). *Sociolinguistic theory: Linguistic variation and its social significance*. Oxford: Blackwell.
- Chambers, J. K., & Heisler, T. (1999). Dialect topography of Quebec City English. *Canadian Journal of Linguistics*, 44, 23-38.
- Chambers, J.K., & Trudgill, P. (1980). *Dialectology*. Cambridge: Cambridge University Press.
- Clarke, S. (1991). Phonological variation and recent language change in St. John's English. In J. Cheshire (Ed.), *English around the world: Sociolinguistic perspectives*. (pp. 108-122). Cambridge: Cambridge University Press.
- Coupland, N., & Jaworski, A. (Eds.) (1997). *Sociolinguistics: A reader and coursebook*. Houndsmills: Macmillan.
- Edwards, V. (1986). *Language in a Black community*. Clevedon: Multilingual Matters.
- Gardner-Chloros, P. (1997). Code-switching: Language selection in three Strasburg department stores. In N. Coupland & A. Jaworski (Eds.), *Sociolinguistics: A reader and coursebook*. (pp. 361-375). New York: Macmillan.

- Gardner-Chloros, P. (1991). *Language selection and switching in Strasburg*. Oxford: Clarendon Press.
- Gordon, E., & Deverson, T. (1998). *New Zealand English and English in New Zealand*. Auckland: New House.
- Gordon, E., & Maclagan, M.A. (1990). A longitudinal study of the 'ear/air' contrast in New Zealand speech. In J. Holmes & A. Bell. (Eds.), *New Zealand ways of speaking English*. (pp. 129-148). Clevedon: Multilingual Matters.
- Herold, R. (1997). Solving the actuation problem: Merger and immigration in Eastern Pennsylvania. *Language Variation and Change*, 9, 165-189.
- Holmes, J. & Bell, A. (1988). Learning by experience: Notes for New Zealand social dialectologists. *Te Reo*, 31, 19-49.
- Holmes, J., Bell, A., & Boyce, M. (1991). Variation and change in New Zealand English: A social dialect investigation. Project report for Social Sciences Committee of the Foundation for Research, Science, and Technology. Wellington: Victoria University.
- Horvath, B. (1985). *Variation in Australian English*. Cambridge: Cambridge University Press.
- Horvath, B., & Horvath, R. J. (2000). Place and apparent time: interpreting the effects of place and space. Presentation at NWA 29, Michigan State University.
- Kontra, M. (1995). On current research in spoken Hungarian. *International Journal of the Sociology of Language*, 111, 9-20.
- Kurath, H. (1972). *Studies in area linguistics*. Bloomington: Indiana University Press.
- Labov, W. (1994). *Principles of linguistic change: Internal factors*. Oxford: Blackwell.
- Labov, W. (1984). Research methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer, (Eds.), *Language in use: Readings in sociolinguistics*. (pp. 28-53). Englewood Cliffs: Prentice Hall.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, D.C.: Centre for Applied Linguistics.
- Labov, W., Ash, S. & Boberg, C. (forthcoming). *Atlas of North American English: Phonetics, phonology and sound change*. The Hague: Mouton. Available: <http://www.ling.upenn.edu/phonoatlas.html>
- Lawson-Sako, S., & Sachdev, I. (1996). Ethnolinguistic communication in Tunisian streets: Convergence and divergence. In Y. Suleiman (Ed.), *Language and identity in the Middle East and North Africa*. (pp. 61-79). Surrey: Curzon Press.

McRobbie-Utasi, Z., & Starks, D. (1999). Variation and accommodation in the individual. Paper presented at NWAV29. Toronto.