
METHODOLOGY IN RESEARCH ON SPEECH PATTERNS: SOME CONSIDERATIONS

NICOLE WONG

Faculty of Languages and Linguistics
University of Malaya

INTRODUCTION

Methodology is a term that is used to refer basically to methods involved in testing a hypothesis or finding answers to a research question. From the outset, the aims of the research and the kind of research govern the methods that are to be applied. It is pertinent to consider, after having chosen a research topic or area and formed a research question, the kind of research to be undertaken. Is it exploratory, predictive, descriptive or explanatory (Dane, 1990)? Once this has been determined, setting the scene is another important consideration (Scholfield, 1995). One has to think of the who, where, when and how of the research. In this article I will discuss the 'how' of research, that is, the research methodology I will look at sampling, the research design, and methods which include data collection, measurement of data and analysis of data.

SAMPLING

Once a hypothesis or research question has been formed, some considerations have to be made on the sampling of data.

Rationale

The rationale for data sampling is closely related to the hypothesis or research question. The rationale has to be established in terms of the following:

- (a) criteria for selection of respondents;
- (b) background of the respondents . academic, social, ethnic, etc.;
- (c) age;
- (d) sex;
- (e) number of samples: what is the 'magic' number?

Ways of Sampling

One has to consider how best to go about collecting one's samples. Two possible ways to think of would be random sampling and systematic sampling.

- (a) **Random sampling** refers to the collection of data at different occasions or at unappointed times where there is no consistent time lapse nor any semblance of regularity. An example is the spread sample of say, 6 or 12 recordings of a child's speech at different occasions which gives a better picture of naturally occurring features in the child's use of language. Observations made of human behaviour in different settings would also constitute random sampling.
- (b) **Systematic sampling** refers to the collection of data that is planned to take place at regular time intervals or at specific occasions or situations which have to be identified from the beginning. An example is the time-sampling or interval-sampling of a child's speech, say from a certain period of time in the morning such as recording ten minutes of the child's discourse every half hour from 9am to 12 noon. The recordings made of a person's oral discourse conducted only in formal situations, for instance, is another example of systematic sampling.

Place and Time

A pertinent consideration for sampling is the question of where the sampling is to be conducted. Would it be indoors, outdoors, local or abroad? A time frame then is crucial for the various stages of the research.

The Variables

The kinds of variables generally considered by researchers are:

- (a) independent or explanatory variable;
- (b) dependent variable; and
- (c) controlled variable.

The independent variable is the variable which gives the researcher more information on the dependent variable. For example, in a study on the speech patterns of native and non-native speakers of L2, the native and non-native status of speakers would constitute the explanatory variable while the average run lengths of the respective speakers would be the dependent variable. The average run lengths would depend on the racial status of the speakers. In other words, the status explains why the average run lengths are higher or lower and which in turn reflect fluency.

Explanatory variables can, therefore, include background of speakers, gender, age and nature of task activities such as formal and informal situations.

Dependent variables, for example, will take into account temporal variables such as pitch, intonation, rhythm, pause length, pause cause, fundamental frequencies, etc.

Controlled variables are those variables which the researcher imposes to exercise control in the research so as to focus on certain aspects of the experiment or study without the analysis, results and findings being affected by other factors. For instance, in the study of speech patterns of native and non-native speakers of L2, the researcher may wish to concentrate on one particular ethnic group of non-native speakers and not on a national spectrum of different ethnic groups because it can be argued that as each ethnic group has its peculiar system of speech pattern, it is not possible to arrive at a national over-

view of speech pattern without compromising the validity of the study of each ethnic group. So the researcher decides to study the speech patterns of L2 of one particular ethnic group of non-native speakers. By doing so, the researcher does not risk encountering what may be divergent results from the speech patterns of other ethnic groups which in turn may interfere with the final analysis and findings.

Ethics

Some thought should be given to the means by which respondents are approached to take part in the research. The matter of personal freedom of choice to participate is a worthy consideration if one is to gather data without being accused of resorting to coercion through promise of monetary payment or gift. Coercion deprives a potential respondent of the choice of saying 'no' at the first instance. The temptation of gains may compromise the quality or degree of willingness of the participation. Similarly, acts of infiltration into teenage groups or some social classes have been deemed unethical since disguises and impersonations are tantamount to acts of betrayal. In the data gathering of language studies, it has been suggested that potential respondents' permission should be sought before participation or given if, for instance, recordings have been made earlier on without respondents' prior knowledge. There are exceptions to existing situations and types of research. In studies on social behaviour, it has been known that researchers do assume certain roles, say, for instance, a homosexual so as to be able to gather information hitherto unavailable to outsiders. However, while such an action may be considered crucial to data gathering perhaps it can only be justified by the strict confidentiality of respondents' identities.

As a rule, the researcher is obliged to let respondents know what the research is about and what is expected of respondents. However, while setting the scene for respondents may help in getting them to participate, informing them too much too early may compromise their participation. It may, for example, result in their being conscious of certain speech characteristics during the actual recording. So the question of how much to reveal at which stage is a delicate one which the researcher may have to figure out as he/she goes along in the data gathering process.

RESEARCH DESIGN

The research design is a way of putting things into perspective. It recapitulates the following:

- (a) the number of variables in the research.

A design with two variables is 'bivariate' while one with more than two variables is 'multivariate' or 'factorial'

- (b) the kinds of variables in the research.

The design explains the kinds of variables involved — explanatory, dependent and/or controlled variables and the relationship to one another, in other words, in relation to the hypothesis.

- (c) what the variables are.

The variables are specified and defined in relation to the research question.

The research design explains how the above (a - c) are all related to the research question or hypothesis.

METHODS

Data Collection

Task Activities

The researcher has to decide what kinds of task activities are required for the respondents. For example, if a researcher is studying the intonation pattern of questions in oral discourse, what kind of activities would be appropriate? Would the activities be contrived and controlled ones? Or should they be natural? Should recordings be made in formal, informal, and spontaneous situations? What would be natural and artificial? What are the definitions and limitations? Some consideration should perhaps be given to the fact that a compromise may have to be made between the natural and the contrived towards a semi-natural approach. For example, it may be ideal for some recordings to be carried out in a natural setting but if in the eventual data analysis, the interferences and background noise picked up in the recording will affect the results, then one may have to consider conducting the recordings in a simulated situation at the studio.

Number of activities and time factor

The next consideration is how many activities should each respondent be required to do. A time frame for the activities is another important consideration as a prolonged session may tire respondents and quality of performance will be forsaken. Enough time should be given to respondents to settle down and do the tasks given without having to rush through the activities and thus compromise performance. If recordings are to be made in as natural a setting as possible, the researcher should allot ample time to make allowances for unexpected occurrences such as technical problems (batteries running out) or early commencement of discourse.

Sampling

One has to be very clear as to what one is collecting and looking for. Are the cases for study the characteristics of the speech of dyslexics or are the cases the dyslexics themselves? Again this consideration harks back to the research question and its relation to the variables. Are the cases observable and measurable? Are they quantifiable (using scores, scales, percentages, frequency counting, coding, etc)?

Number of respondents

Researchers often wonder how many respondents would be an ideal number. Would five be too few or would ten be the magic number? Is it best to collect data from as many respondents as possible? One has to consider the number of recording sessions, duration of each session, number of activities and situations that each respondent is subjected to and that may give one a fairly good idea of how many respondents to approach. Thought should also be given to the possibility that some data may not be usable due to some technical faults or excessive noise or interference during recordings, in which case, having a few extra respondents may be deemed a wise move. It is no good either to collect data on the premise that more is best because ultimately one has to consider how the data is to be handled, organized and analysed. In the collection of raw speech data, transcription of recordings has been widely known to be an arduous task. An hour's recording may take three to five days' work of transcribing and typing.

Instruments/tools required

Is the data gathering to be conducted indoors for example, in a studio or outdoors? The instruments required may differ according to the surroundings. If the recording is to be carried out in the audio studio, then bookings have to be made beforehand and according to the convenience of respondents. Needless to say, instruments should be of good quality and durability. It is also wise not to overlook simple things like having new batteries on standby when old ones run out or when there is interruption to the electrical power supply. The researcher should set up the paraphernalia, be ready before the respondent arrives to ensure that nothing is amiss and that the recording will proceed smoothly without undue waste of time and needless stress to the respondent.

In data collection, it cannot be emphasised enough that one has to be very clear about what one is looking for, about the amount of time spent on collection, transcription, and quantification, and how the data is to be observed and measured. It is very easy to be sidetracked, for example, to confuse reading skill with oral skill in spontaneous speech. It is tempting to collect more and more data that becomes unwieldy and too time-consuming for classifying and analysis, or conversely, to collect insufficient data. One has to be prepared that some data may have to be disregarded due to technical faults such as recordings interrupted by noise, coughing, power failure, etc.

An important consideration is to carry out a piloting of the data gathering as this will enable one to improve on one's research design and methods and upgrade the face validity of the research methods.

Measurement of Data

The measurement of data is closely related to the cases. Here one has to consider how one goes about measuring what is observed. For example, in studying fluency one observes the average run length, in research on pitch there are computer software programmes that measure the fundamental frequency or show the intonation pattern graphically. It is important for the researcher to consider the availability of such tools of measurement. Also, one requires training or some rudimentary technical knowledge in the use of the necessary tools in the area of phonetics and speech such as instruments for converting speech signals to observable spectrograms and ways of measuring aspects of prosodies e.g. VOT (voice onset time) for certain consonants.

There may not be a tailor-made computer software that is ideally suited to one's needs and therefore, one has to be prepared to slouch over the computer at the speech laboratory for many hours of tedious work. A paragraph of data may require a couple of days' work.

Analysis of Data

Now that all the data has been gathered, transcribed and organised, how does one carry out the analysis? In the statistical analysis of data, some tools of analysis are computer programmes like the SPSS programme, miscue analysis such as the quantifying of slips of the tongue, hesitations and errors in reading aloud (Scholfield, 1995), and chi-square tests which compare earlier data to current data or which compare data of subsamples (Dane, 1990), to mention a few. For some of these tools of analysis, prior training may be required and one may have to attend lectures or courses. Another alternative would be to employ an analyst to assist. One also needs to consider the accessibility of such tools as computer programmes.

One has to be prepared to spend a lot of time, sometimes months, looking for a programme or tool and learning how to use it. There will be hours spent staring for days (and months) at the computer terminal in the analysis of data.

There is also the question of what one does with data that is inconsistent or discrepant. Does one ignore it or include it? What do the results tell about the research, in particular, the hypothesis?

CONCLUSION

The method is the 'how' of the 'what' of the research, i.e. the research question. One may well come up with a brave topic, but in the final analysis, one has to be able to collect, observe and measure the data and have the means of quantifying it. What follows is the quantitative analysis of data and interpretation of the results that may strengthen the hypothesis or require some modifications to be made to it.

