

Creating a Corpus of Linguistic Field Data

A Focus on Eclecticism

By: Ammaria, Derni
University- Tlemcen

Abstract:

In response to the vast technological progress and to the increasing issues and tools in the field of research making, it is of importance to explore various techniques and instruments to create and annotate a reliable corpus of linguistic data. The present paper raises how the combination of different procedures and techniques can increase the flow of data, and thus, help researchers to improve the efficiency of their work studies. It reports, despite the drawbacks of the eclectic approach, examples of mixed method techniques aiming at expanding the scope of qualitative and quantitative research in the field of linguistic science.

Content:

1. Purpose and Assumptions
2. Eclecticism
3. Qualitative and Quantitative Methods
4. Data Collection
 - 4.1. Observational Data
 - 4.1.1. Recording Spoken Interactional Data
 - 4.1.2. Discourse Analysis
 - 4.2. Elicited Data
 - 4.2.1. Interviews
 - 4.2.2 Questionnaires
5. Discussion

1. Purpose and Assumptions:

Language may be speculated from at least five angles, accordingly linguistic, biological, psychological, social, and ecological. In the linguistic approach, the structures of a system are identified, described and explained at different levels of analysis namely, phonology, morphology, syntax and semantics. The linguist's aim is to analyse language with an explicit attention to the linguistic features as they are structured in the system.

The biological procedure rises up questions that relate language with biology. It determines biological factors that make humans able to acquire and

use a verbal system of communication. In this way, biological preconditions for language use are taken into account.

In the psychological approach, linguistic behaviour is studied in relation to the psychological processes that underlie it. Language is therefore regarded as not merely a matter of verbal choices, but also as a product of interpersonal and intrapersonal motives. Speakers are thus considered as individuals of different linguistic performance abilities. The correlation between linguistic behaviour and the human psyche is one concern of researchers. The other direction includes the role of language as it effects memory, perception, recognition, thinking and learning.

The social procedure points out how social factors come to control and contribute language use in a given context. Linguistic differences correlate with a number of social variables. Age and gender for instance are the most common attributes to the choice of some linguistic forms at the expense of others. Social realities in this way do specify conditions under which some choices are appropriate more than any other alternatives.

The ecological approach supports the view that language is used in unpredictable and unconstrained ways. Ecolinguistic criteria like the affiliation of language, the identity of the speaker and the ecological structure he belongs to as a member of a particular speech community, living in a given environment characterised by identical social and cultural values and governed by an ecological milieu, do bring to light the conditions under which linguistic forms are used. Ecolinguistics is therefore the study of language according to the environment it is used in. The study of natural ecological phenomena in correlation with language is also the concern of the ecological approach.

The fact that language enters into almost every aspect of human life makes the application of linguistic analysis of a broad type. The present paper deals with human language as it is viewed by the above approaches in the field of linguistic analysis. Whatever is the base under which a system is tackled, linguistic research includes different facets of language from the articulation and the description of speech sounds to the speakers' attentions to converge or to diverge. Besides, it is widely viewed that any empirical study requires first and foremost reliable data to work on. Yet, despite the wide range of techniques that are held nowadays by researchers in the field of data collection, linguists always query the reliability and the validity of the corpora used to conduct research. The aim of this article is to show how the combination of different procedures and techniques can increase the flow of data. Eclecticism which is the concern of this paper is viewed from mainly two different prospects: data collection and data analysis.

2. Eclecticism:

Crystal affirms that:

“ Eclectic accounts are justified by the multiple insights they can provide into the area of language.”

(Crystal:

1983, 125).

Apparently, eclecticism is an approach that does not hold a single paradigm i.e. a set of assumptions. It draws upon multiple methods and theoretical frameworks. It is therefore the process of selecting and compiling features from different sources.

Eclecticism may either be at the paradigm level, at the method level, or at the technique level. The weakness of the combination of different paradigms into a single inquiry leads to *“ the difficulty of developing a coherent theoretical framework within which the various descriptive components can be interrelated”* (Crystal: 1983, 125). Because linguistic approaches entail distinct contradictory positions, it is not obvious to accumulate them. Otherwise, the linguist is at the risk of mixing up everything. The intention of this paper is to bring the audience to an understanding of eclecticism at the methodological level, as far as the combination of quantitative and qualitative research methods are concerned with a particular emphasis on data collection and data analysis techniques, as they are not linked distinctively to a specific linguistic approach. No technique is typical to a particular approach in separation with others either the structuralist, the functionalist, or the generative school for instance. The eclectic approach is used with a particular focus on:

1. the combination of qualitative and quantitative research methods.
2. the combination of observational and elicited data.

3. Quantitative and Qualitative Methods:

“Qualitative methods of research” is a general term describing research practices associated with a variety of field inquiries. In linguistics, they aim at studying language as it is used in naturally occurring spoken and written forms. They are therefore concerned with a naturalistic and an uncontrolled observation of language behaviour, whereas quantitative methodologies rely on experiments to test a specific hypothesis using statistical techniques to describe some aspects of language use through numbers and graphs. In qualitative studies, results are interpreted in words through the comments and the remarks the researcher comes up with.

The idea of an absolute objectivity with qualitative investigations is somehow rejected, as the investigator interprets a linguistic behaviour according to the observations he makes in identifying linguistic patterns as they occur in language. Contextual factors are analysed and explained vis à vis the findings he gets. The problem with such a procedure is that it lacks objectivity. Findings may differ from one researcher to another. Differences of this type may be reduced with a good practice in conducting research and through making valid interpretations based on a suitable theoretical framework.

Qualitative sampling is also subject to individual variation. Such a variation is to be reduced if not eliminated in quantitative investigations. This kind of research supplies researchers with narrow data, as the number of participants is much more limited in comparison with the natural use of language, which involves the whole society or the whole speech community.

The problem with quantitative studies is that it is not easy to find informants who fit the hypothesis a linguist is arising in his work best, whereas the naturalistic observations in qualitative investigations allow linguists to explore and analyse linguistic behaviour in real contexts. Yet, it is more difficult to control linguistic facts in unlimited contexts than in guided fieldworks.

The discussion above facilitates an answer to the question: "To what extent can a researcher complement quantitative with qualitative methods - and vice versa - within the same research design?" The comparison of the two research methods across their advantages and their drawbacks suggests that both empirical methodologies may be used in a complementary way, so that the weaknesses of the qualitative procedures are remedied by the advantages of the quantitative procedures and vice versa. The researcher therefore has the opportunity to promote the validity of both procedures through handing on a single research design with reference to both of them. Qualitative and quantities methods used in conjunction may provide complementary data sets which together give a more complete picture than can be obtained using either method separately.

4. Data Collection:

It is widely argued that any empirical study of linguistic behaviour is based on linguistic data which constitute the subject matter of enquiry. Investigators are likely to use different techniques and tools at this phase of research. Corpus data have always been used as an empirical basis for describing and determining linguistic systems at different levels of analysis and in all fields of enquiry. It is of importance to constrain and to understand the areas of interest of an investigation before conducting data collection. The researcher needs to

focus on a specific linguistic form to work on. . It is also important to bear in mind that a good sample contains a great deal of relevant data. The following are some techniques of getting a corpus of reliable linguistic data:

4.1. Observational Data:

Any kind of data acquired through observation through the eyes and the ears of the investigator are included here. It is not necessary to the researcher to take part in the linguistic behaviour. Two major techniques are to be taken into account. These are:

4.1.1 Recording Spoken Interactional Data:

Where any kind of audio or video recorder can be used. The recordings may be kinds of ordinary conversations between friends, members of the same family, relatives, colleagues...in a company, a store, a school, or elsewhere. Not to get into an embarrassing situation and not to influence the participants, the researcher has to hide carefully the recorder without impeding the procedure. Such a technique raises moral issues, as the privacy of the recorded participant is to be respected. The investigator has to be careful if there are any personal or confidential information that are revealed in the recording not to hurt the speakers involved in such natural conversational data. If he does not feel the need to participate in conversations for one reason or another, he may dispense with. The recorded samples may be digitalized in a digital recorder, or downsized under the form of sound tracks on a computer. Such a procedure helps the investigator to analyse the data and to compare his findings.

4.1.2 Discourse Analysis:

A discourse is any type of composed language whereby speakers create sentences, paragraphs, essays and the like. Both the spoken form and the written form are included here. Discourse analysis is concerned with the production and the interpretation of language in situated contexts in relation to social and cultural norms. It can proceed at various levels. Attention may be focussed on a lower level where individual units such as lexemes and phrases are taken into account, or at a higher level where larger units such as whole sentences, paragraphs and essays are analysed. The investigator must take into account the triad of the sender, the message, and the receiver into consideration. Phonological, morphological, syntactic and semantic criteria are also explored here. This technique changes according to the context or the kind of discourse worked on. Sermons for instance are not analysed in the same way commercial advertisements are.

4.2. Elicited Data:

Including any kind of data got in response to questions asked by the investigator himself. Many elicitation techniques can be held in field works. The ones we are considered with here are:

4.2.1. Interviews:

Where the investigator himself participates through giving questions seeking at reaching the linguistic forms he has fixed as a goal in mind. Responses are open-ended and can be recorded by written note taking or on a recording device. Informants are asked to discuss their lives and their interests. They are pushed to tell stories so that they become engaged in the subject matter of the interview. They can be encouraged to keep talking by the researcher's repetition of their utterances. He can also show interest through smiling, head nodding and eye contact. Pressure is relieved from participants by reassuring them that there are no correct or incorrect responses to any issue which may be raised during the conduction of the interview. Such a technique can also be held in groups of speakers where a more natural interaction is likely to occur than with a formal face to face interview. Toward the end of the conversation, more elicitation devices such as word lists, questionnaires, minimal pair lists, drawings...may be added if appropriate. These interviews may be recorded and digitalized too.

4.2.2 Questionnaires:

Questionnaires are primarily document based. Accordingly, they lack context. With the use of this technique, the emphasis is on quantitative data more than qualitative data. Their design is a bit hard and time consuming, as they require an initial knowledge of domain. Besides, the investigator has to be very careful to select the questions and the items he has to go through. This enables him to elicit the right kind of knowledge he requires in his fieldwork. Questionnaires can also be useful to compare between different groups.

5. Discussion:

One general difficulty with data collection in observational techniques concerns the spread of domain knowledge as the linguistic forms the investigator aims at studying may be distributed across different sources. Such data are scarcely if not never found under an explicit form. They are not written down. Besides, there may be variations in these various contexts. Within the use of elicitation devices, native speakers are not aware of the language system they acquire in their infancy, that is why they find it difficult to describe mere forms of speech they usually come through in their daily life. The presence of the investigator might also effect the informants. They may not feel free to tell the

researcher what he needs to know. They may also hide information on purpose. They can supply the investigator with superficial responses, or provide him with falsified data. The problem with observation is that it is time consuming, as it requires much more time than with elicitation techniques. However, it permits the investigator to deal with details that other techniques cannot afford. Such social contextualised and rich data are of course hard to analyse.

Observational data are all qualitative, since they are descriptive and not based on numbers and graphs. Elicited data have a limited observation. They are manageable, and it is easier to analyse them. They are carefully selected and not acquired at random.

Because neither the observational nor the elicited data are perfect, they can be used unmistakably in combination. The complementary use of the various techniques allows the investigator to cover the weaknesses of each. The genius in linking the elicitation devices with the traditional observation procedures will inevitably show how skilful is the investigator and how eclectic can he be in conducting scientific research in linguistics. The investigator has also the opportunity to quantify what is qualitative and qualify what is quantitative.

References:

- Achard, P. (1998) La Sociologie du Langage Presse Universitaire de France: Paris
- Bell, R.T. (1976) Sociolinguistics: Goals, Approaches and Problems, B.T Batsford Ltd: London
- Chambers, J.K. and Trudgill, P. (1980) Dialectology, Cambridge University Press.
- Crystal, D. (1983) A First Dictionary of Linguistics and Phonetics Andre Deutsch: London
- Hudson, R.A. (1996) Sociolinguistics Rev.ed, Cambridge University Press. Great Britain
- Hymes, D. (1964) Language in Culture and Society: A Reader in Linguistics and Anthropology, New York: Harper and Row. Press. Great Britain
- Lyons, J. (1986) Language and Linguistics, Cambridge University Press. London
- Spolsky, B. (1998) Sociolinguistics Oxford University Press. United Kingdom
- Widdowson, H.G. (2000) Linguistics, Oxford University Press. New York
- Web Sites:
- Connell, J., Lynch, C., & Waring, P. (2001, December). "Constraints, compromises and choice: Comparing three qualitative research studies". *The Qualitative Report*, 6(4). Retrieved [Insert date here], from <http://www.nova.edu/ssss/QR/QR6-4/connell.html>
- Fettes, M (1997) "Toward an Integrative Linguistic Science: Stabilizing What? An Ecological Approach to Language Renewal" A revised and expanded version of a paper presented at the Fourth Stabilizing Indigenous Languages Symposium, Flagstaff, Arizona, from [http://www.esperantic.org/~mfettes/linguistic ecology toward an Integrative linguistic science stabilizing what /sil.htm](http://www.esperantic.org/~mfettes/linguistic_ecology_toward_an_Integrative_linguistic_science_stabilizing_what/sil.htm)
- Roberts, A. (2002, September). "A principled complementarity of method: In defence of methodological eclecticism and the qualitative-quantitative debate". *The Qualitative Report*, 7(3). Retrieved [Insert date here], from <http://www.nova.edu/ssss/QR/QR7-3/roberts.html>
- Sandelowski, M (2000). "Combining Qualitative and Quantitative Sampling, Data Collection, and Analysis Techniques in Mixed-Method Studies", from [http:// www.phac-aspc.gc.ca/publicat/cdic-mcc/18-3/d_e.html](http://www.phac-aspc.gc.ca/publicat/cdic-mcc/18-3/d_e.html)