10-2008

LANGUAGE A PHENOMENON OF CULTURE AND COMMUNICATION, PRESERVING THE LANGUAGE THROUGH A PHONETIC MAP

Ben Abdessalem Wahiba

Institut Superieur de Gestion de Tunis, Tunisie, wahiba.abdessalem@isg.rnu.tn

Follow this and additional works at: http://aisel.aisnet.org/mcis2008

LANGUAGE A PHENOMENON OF CULTURE AND COMMUNICATION, PRESERVING THE LANGUAGE THROUGH A PHONETIC MAP

Ben Abdessalem, Wahiba, Institut Superieur de Gestion de Tunis, 41 Rue de la Liberté, Cité Bouchoucha, 2000 Le Bardo, Tunis, TUNISIE, wahiba.abdessalem@isg.rnu.tn, wahiba.time@yahoo.fr

Abstract

The language is the key instrument by which we assimilate the culture of our country. As culture is influenced by various factors related to history, traditions, etc. language is also affected by these phenomena. Their influence may differ from one region to another; this can produce specific regional language known as dialects.

The dialect encloses lexical, syntactic or phonetic particularities of the language, used by a part of the population or some regions of the nation.

The Tunisian dialect is becoming increasingly important: it is used more and more by Tunisian chains of TV, newspapers, etc. However, several expressions have become unusual for some people, some regions or some generations. This phenomenon is due to the influence of the language by several social and cultural factors. These dialects deserve to be studied in order to preserve them and to keep them in memory.

This paper consists of two, disparate parts. In the first part, we present the experiences of many years of fieldwork made by the staff of linguistic Atlas in Tunisia. As a result of the fieldwork, the project now has access to recordings from approximately 3000 speakers, describing the Tunisian dialect, in more than 250 Tunisian regions. We are currently pointing research on various aspects of the sound patterns of these dialects: the phonetic system. The second part of the paper is meant to give an overview of an automated system describing the phonetic system of the Tunisian dialect. This system can be very useful in understanding the variation of dialects in different geographical regions of Tunisia through an intelligent map representing variations of sounds as we move from one place to another.

Keywords: culture, dialect, phonetic map, natural language processing.

1 INTRODUCTION

Our work falls within the framework of activity of CERES (CEntre de Recherches Economiques et Sociales). The centre was placed administratively under the State's Secretary of Scientific Research and Technology.

Its task is to perform all research activities and studies in the field of economics, social and human sciences. In this setting, it undertakes investigations and surveys for identification, analysis and handling in the past or in the present of social and economic phenomena, in order to proceed to all anticipations and prospective behaviour (Ceres, 2003).

The language is closely related to culture but also to several phenomena of society (Kaye, Alan S. 2006). In this context, CERES has appealed to numerous collaborators of different specialities: linguists, computer scientists, etc. to develop a system related to a linguistic Atlas of Tunisia.

The linguistic Atlas is a linguistic investigation fieldwork aimed at collecting data in order to describe the Tunisian dialect system.

This effort will be more interesting if it is be associated with a talking Atlas: listen to recording that can complement linguistics descriptions of languages (lexical, syntactic or phonetic).

In this paper we are interested only to the phonetic aspect of the Tunisian dialect. The main purpose of the system is the description of a computerized system that allows scanning, storing, manipulating and representing data on the phonetic system of the Tunisian dialect.

Our paper is organized as the following: In the first section we describe the dialect as a socio-cultural phenomenon. In the second section, we present the approach and tools used for the realization of such a system. The third section is a conclusion of this work.

2 DIALECT AND SOCIO-CULTURAL PHENOMENA

The dialects of the Magreb (Algeria, Morocco, Libya, Tunisia, etc.) were influenced by historical events (Caubet, D. 2001). Thus, the Tunisian dialect records words or expressions borrowed from different cultures.

In fact, relating to the history of Tunisia and its numerous invasions, we mention that the Tunisian dialect, integrated the Berber, Roman, Arabic, Turkish, French vocabulary, etc. In the Tunisian dialect, we note that there are several heritages. Any foreign term responding to the need for expression or denomination is naturally integrated into the lexicon. Many words of French origin were included, such as: [Sminijja], "cheminée" (chimney), [tri:si :ti], "électricité" (ectricity), [tnerviz], "s'énerver" (to be nervous) (Mejri, S.2000),or of Italian language: [kugina], "cucina" (kitchen) or English language: [skula], "school"

This shows that the dialect is not a simple tool of conversation and expression, but it is also, the support of a culture and a vision of the outside world.

The dialect can be also seen as a social phenomenon. Indeed, in the Arab world, for example, dialect studies distinguish generally between nomadic and sedentary dialects and, within that second category, between rural and urban dialects (Miller, C. 2005). Namely, the fact of coming to live in the city and becoming a citizen, usually involves the adoption of talk of the town.

Other social phenomena are correlated with the dialect: age, sex, level of education, social status, etc. For example, many words are used only by a part of the population: youth or older, men or women, rich or poor, and so on (Vanhove, M. 2000).

3 THE APPROACH

To realize a linguistic Atlas this requires first and foremost a linguistic fieldwork: *questionnaire*, which will serve as raw material on which linguists base their studies and syntheses. The questionnaires were made in the form of interviews that have been recorded. These recordings can be well exploited by computer programs.

3.1 linguistic data collection:

The realization of sound recordings with those of the studied population is a reliable method, since everything said is recorded and can be listened and used several times.

For the sake of the linguistic Atlas of Tunisia we collected data on the Tunisian dialect, it was necessary to cover 250 regions in the Tunisian territory, to carry out an investigation in each of them. The investigation implies four kinds of people called informants (two young people: a boy and a girl and two adults: a man and a woman). The investigation is composed of three types of

questionnaires: phonetic, syntactic and morpho-syntactic. The questionnaires are organized into chapters; each chapter is composed of a set of questions. The recording session of each survey is about one hour by questionnaire. The result of this work is a set of more than three thousand audiotapes. To ensure proper monitoring of these recordings, to each recording session corresponds three documents:

An investigation sheet:
This gathers information on the investigation such as: region, date, time of the investigation, name of the investigator and that of the informant, and so on.

An informant sheet:
This contains information about the informant: identity, residence, instructive level, social environment, and so on.

An investigator sheet:
That is a set of information related to the interviewer: identity, qualifications, number of surveys, and so on.

A registration sheet:
it includes information such as code of registration, rank of the investigation, name of the informant, name of the investigator, date of the survey, date deposit, etc..

The information contained in these documents allow firstly to uncover the actors of the investigation (informant and investigator) and secondly, to clarify the spatio-temporal framework of the investigation.

Under a phonetic questionnaire, the information required concerns the meaning or the vocalic matter (Baccouche, T. and Mejri, S. 1998). The questionnaire deals with phonetic chapters such as:

✓ The opposition ق [q] / ڤ [g]: in the Tunisian dialect opposition [g] / [q] has always been regarded as the transposition of phonetic character between the Bedouin and the urban dialects. As in the word relative to the *blue* colour which can be pronounced in two manners, with [g]: زرڤة [zerga] or [q]: زرقة [zerqa].

✓ The attack glottal: (hamza), which involves a glottal vibration. Example the words: "Koran": قرآن [qor3an] \ قرعان [qoraan] or "question": سهل [shal] / [s3al] سأل. The first word is used mostly by old people,

✓ The emphasis: check if the emphasis is a phonetic variation or reflection of a change of meaning. The change of vowel ɛ/a causes an emphasis. Example of emphasis the word: حار can be pronounced [ha: r] or [hɛ: r] and imply either the verb *"hesitate"* or the adjective *"spicy"*.

✓ Aperture Intermediate before diphthong: is a unique vowel that the tone changes at pronouncing. Example: the word *"where"* that can be pronounced in three ways: [fi: n] in some regions, [fe: n] or [fajn] in other regions.

3.2   Handling of sound recordings

As the data used are sound recordings on magnetic tapes, so in analogical format, it was necessary to carry out some operations for example: the transformation of these recordings from the analogical format to the digital format, the segmentation of the recordings, and so on.

3.2.1 The switch to digital format

To make sound recordings used by computer programs, we had to make a choice among ranges of software tools to make a transition from analogical signals to digital signals. In addition, we had to take into account several constraints as size and quality of the result.

A large number of software techniques have been used for the treatment of sound recordings. *Jet Audio*, was the most convivial and easier to use for the conversion from the analogical format to the

digital format. The result of this operation is a *WAV* file, which has not a good quality. *Sound Forge* has been used for cleaning these files and improving the sound quality: reduction of breath and cracking (or noise). As the tapes had a speed twice as fast as the normal speed, it was necessary to use the software *MIXVIBES* to normalize the speed of the sound. To minimize the storage space, a compression sound files was needed. We tried towing kinds of compressions: a non-destructive compression and a destructive compression.

    ✓ The non-destructive compression:

The main property of non-destructive algorithms is that once decompressed, the file is exactly the same as the file before compression (the digit near). There is therefore no loss of information. This is the case with *zip, arj, rar c*ompressions. However, this type of compression does not give an interesting compression rate for the sound (CUEFF, A.1999-2002).

    ✓ The destructive compression:

This compression favours the compression ratio. However, it is accomplished by removing some parts of the audio signal, some frequencies supposed to be hidden by others (Psycom, 2002). It is a compromise between compression ratio and quality. There are many destructive compression formats: *MP1, MP2, Real Audio,* the most famous *MP3*, etc. This is the *MP3* format that we used for our system; it is the best according to quality and storage space. Numerous software tools can convert an audio file from one format to MP3 format, for example: *dBpowerAMP, Exact Audio Copy, Adobe Audition, MP3 converte*r, etc. It is the latter that we have chosen, it is very easy to use.

3.2.2 The Record segmentation

The recordings contain a whole conversation with the informant to push him/her to pronounce specific words spontaneously. Hence, it was necessary in a second stage, to treat these recordings to be able to extract the records fragments corresponding to the words addressed by the questionnaire. The software *Sound Forge* allows this kind of process.

3.3   Realization of the database

The database can store pieces of formations related to informants, investigators, and regions of investigation, investigation, questionnaires, chapters and questions. The only difficulty is in the *question* table which includes mainly the field word, which is a word written in Arabic characters and phonetic transcription of the word. This has necessitated the use of Latin and Arabic characters and other symbols to reflect the specificities of the Tunisian dialect.

3.4   Map treatment

The main goal of our system is to listen to a given word by an informant in a region selected on the map of Tunisia. A set of information can be visualized to describe the word (information on the informant age, educational level, social level, Information on the location of the investigation, etc.). The treatment of the card is to achieve a kind of communication between geographic areas on this map, the database and the registration segments corresponding to a chosen word.

Since a geographical parameter is implicated in this type of study, an important cartography work was necessary, especially when representing information on the map. The latter is not static, but changes in behaviour, according to the needs of the user and information stored in the database. The programming language *VISUAL BASIC* allows this communication, but we need in also some *CONTOLE ACTIVEX* developed with *VISUAL C++* programming language.

4   CONCLUSION

This work can be considered as a model of linguistic Atlas of Tunisia in a computerized version. It describes the phonetic system of Tunisian dialects which could be used to describe the variations of phonetics vocabulary through an intelligent map. It is a kind of exploratory tools and techniques used to achieve a complete Atlas describing the specificities of the Tunisian dialect (phonetic, morphological, syntactical, etc.).

Given the continual technological progress, it would be necessary to look for other software tools to have a better sound quality, better compression ratio of audio files, and so on.

However, with the recent progress of information technology, it would be also, possible to offer linguistic information in the audiovisual form (sound and video), which combines the language and gesture, the verbal and the non-verbal (Kristol, A. 1997).

References

Baccouche, T. and Mejri, S., (1998) : L'Atlas linguistique de Tunisie: problématique phonologique, Aida III, Malta.

Caubet, D. (2001). Questionnaire de dialectologie du Magreb. Estudios de diallectologia norteafricana y andalusi, n° 5, 73-92. .

Ceres, (2003). http://www.mes.tn/ceres/fr/pres.html

CUEFF, A., (1999-2002). http://www.oamao.com/Dossiers/Compress/compression.htm

Kaye, Alan S. (2006). Dialect, Culture, and Society in Eastern Arabia, vol 2:Ethnographic Texts.(Book review). The Journal of the American Oriental Society. Kristol, A., (1997). Un nouveau type d'atlas linguistique : l'Atlas linguistique audiovisuel des dialectes franco-provençaux valaisans (ALAVAL). Les Cahiers du CEL, n°1, 141-151.

Mejri, S., (2000). L'Atlas linguistique de Tunisie: le questionnaire morphosyntaxique, Aida IV, Marrakech.

Mejri, S., (2000). Le petit prince en arabe dialectal tunisien, Meta, vol. 45, n° 3, 564-565.

Miller, C., (2005). Les Sa'îdi-s au Caire : accommodation dialectale et construction identitaire. Iremam, Aix en Provence.

Psycom, (2002). http://mao.audiofanzine.com/apprendre/dossiers/index,idossier,31,page,3.html

Vanhove, M., (2000). Processus de koïnisation dans le dialecte arabe de Yafi' (Yémen), Aida2000.