

Amazigh Language Terminology in Morocco or Management of a 'Multidimensional' Variation

Aicha Bouhjar

IRCAM

B.P. 2055 - Rabat

E-mail : bouhjar@ircam.ma

Abstract

The present communication brings to the fore the work undertaken at IRCAM within CAL within the framework of the language planning of Amazigh, particularly on the side of terminology. The focus will be on the concept of "variation" that affects different levels in the course of standardizing a language: orthography, spelling, grammar and lexis. Thus, after a brief survey of the main features of the Amazigh (Berber) language in general, the missions and the projects far achieved by CAL will be presented, particularly the objectives that relate to the work on the multiply varied corpus-based terminology. It appears that eliciting the pertinent information, for the most part, requires a whole amount of work on the "re-writing of corpora so that the latter become exploitable in the standardization process. It should be pointed out that this stage of data homogenization, seemingly unwieldy for optimal exploitation, cannot be undertaken Amazighist linguists being involved in theoretical and methodological presuppositions that are at the root of this variation.

Introduction

Subsequent to the Royal Speech of Ajdir on October 17th, 2001, the Amazigh culture has been endowed with an institution whose missions are, inter alia, the standardization and promotion of the Amazigh language (Berber). This institution is officially known as the Royal Institute of the Amazigh Culture (IRCAM, henceforth).

Such missions, on the whole carried out by the Language Planning Center known as 'Centre de l'Aménagement Linguistique' (CAL), incarnate some long-term objectives that the center intends to achieve. CAL has been operational since 2001, and over a five-year-period, the center has realized a number of important activities, including the codification of the Tifinaghe alphabet, the development of a standard Amazigh grammar and the instigation of an institutionalized gamut of Amazigh terminology. This very point calls our attention most, inasmuch as work on terminology is generally based on the various components of language, including the lexicon, morphology, syntax and sociolinguistics. Since the present work concerns a language on the way to standardization, the case being that of Amazigh in Morocco, a number of challenges have to be faced in order to meet the incessantly increasing terminological demands. The challenges that the present contribution purports to investigate relates to variation that is characterized as 'multidimensional', the latter covering aspects relative to written form, transcription method, regional variation and corpus diversity.

Before dealing with the issue relating to the work on terminology that has been initiated in Morocco, it is of

relevance to give some information about the Amazigh language and to introduce CAL in some detail, emphasizing the progress the center has made in matters concerning the codification of the Tifinaghe alphabet and the elaboration of a standard Amazigh grammar.

2 Short characterization of the Amazigh language

Below, we present a sketchy table that captures the different levels of the Amazigh language.

Thus, though in a geographically discontinuous manner, the Amazigh language stretches over an area that is so vast, from the Canary Islands to the Siwa Oasis in Egypt to the north and from the Mediterranean coast to Niger, Mali and Bukina Fasso to the south. Historically, the Amazigh language has been autochthonous and is a member in the Hamito-Semitic or "Afro-Asiatic" family. On the linguistic side, the language is characterized by the proliferation of dialects due to historical, geographical and sociolinguistic factors. In Morocco, one may distinguish three major dialects: the Tarifiyt in the northern, the Tamazight in the center and the Tashlhiyt in the southern parts of the country. According the recent demolinguistic data, the Amazigh language is spoken by some 30% of the Moroccan population (around 10 Million inhabitants). Politically, the Amazigh language has enjoyed its status at different levels, depending on the country where this language exists. In Morocco, the status of Amazigh has achieved the most advanced level, especially following the Ajdir Speech of 2001, laying the foundation of IRCAM in a Dahir which stipulates that the Amazigh culture is a "national matter" and, this being the case, is a concern of all the citizens (more than 30 million in number).

3 Presenting CAL

The task of CAL is to undertake research on the Amazigh language, aiming at progressive and convergent standardization at the levels of phonology, lexicon and grammar. To meet this end, CAL works in close coordination with the centers concerned with the implementation of the Amazigh language in the educational system, mass media and public life in general. One of the missions of IRCAM being the preservation of the cultural heritage of Amazigh in all its forms, one further task of CAL is to collect data in different language areas of Amazigh. In so doing, the center aims both at preserving the language and the publishing reference works in this domain.

CAL comprises two Research and Study Units (RSU) :

The lexicon RSU (UER-Lexique)

The grammar RSU (UER-Grammaire)

The activities of the CAL span different levels, particularly action research, research by contract, and the training of teachers and teacher trainers in Amazigh.

4 The Rebirth of the Amazigh Alphabet in Morocco

Like any language undergoing a passage from an oral to a written mode, the Amazigh language has been in need of a graphic system, a task which CAL has always prioritized. In this regard, there were three competing graphic systems for Amazigh, all submitted to a scientific and technical study at the center. These competing systems are those of Arabic, widely used by clerics and poets in rural areas, Latin, used particularly by berberists since early works of missionaries, and Tifinaghe, an ancestral writing system which has to date been preserved by Touareg, particularly by women.

In Morocco, the choice ultimately fell on Tifinaghe for technical, historical and symbolic reasons. Since the Royal declaration on February, 11th, 2003, Tifinaghe has become the official graphic system for writing Amazigh, particularly in schools. The scientific and technical study of the Tifinaghe alphabet (Ameur and *al.*, 2006a), which motivated its choice, went through six steps:

1. An exhaustive study of the different variants of Tifinaghe (from rock inscription to neo-tifinaghe fonts).
2. Evaluation of the adequacy level of the different notations based on the analysis of phonological and morpho-syntactic structures of Amazigh.
3. Reliance on such criteria as grapheme univocity (one sound = one grapheme), economy, consistency and historicity.

4. A cost/gain study based on the principle of simplicity on the technico-pedagogical side.
5. Internal piloting (IRCAM's scientific committee) and external piloting (during teacher training sessions and in learners' classrooms) of the proposed graphic and orthographic system.
6. Finalization of the graphic system.

Tifinaghe-IRCAM, which tends to be a phonological alphabet, consists of 33 letters, thereby permitting the writing of the Standard Amazigh, with a margin of possible regional pronunciations. The elaborated alphabet, then, is one that allows written intelligibility nationwide. Thus, the grapheme (letter of the alphabet) X can be pronounced as [g/g/j/y].

Example: ⵙⵓⵍⵓ 'sky' pronounced agnna/ agnna/ aynna/ ajnna

The Tifinaghe-IRCAM graphic system as well as the alphabetic sorting have been officially recognized by the ISO-Unicode (for the graphic system : June 2004) and the SNIMA (Moroccan normalization service).

5 The Standard Amazigh Grammar

Initiation à la langue Amazigh (Introduction to the Amazigh Language), the first work published by CAL, succinctly describes the functioning of the structures of Amazigh in Morocco, account being taken of the diversity of the language. However, working towards a Standard Amazigh language, the book puts particular emphasis on what the Amazigh varieties have in common. A subsequent work ensued which contains a more exhaustive treatment of the Amazigh grammar. Hence, the work accomplished at CAL is not representative of individual dialects of Amazigh, but rather of the totality of its varieties. CAL has placed an all-incompassing and integrative vision in its treatment of grammar, just like the strategy it put forth in the selection of a graphic system to write Standard Amazigh. For example, if ideosyncratic structures arise in given dialects, the decision is to treat them as competing structures at the speaker's disposal.

6 Amazigh Terminology in Morocco

Since the creation of IRCAM, Amazigh terminology has been adopted at the institutional level. In actual fact, the RSU concerned with the lexicon is all too often requested to help respond to specific terminological demands on the part of many public and private organisms. The oft-heard question that the research team is expected to answer is 'how do we say X in Amazigh ?' (X being a term either in French or in Arabic, account being taken of the sociolinguistic profile of Morocco). The terms requested belong to such different

domains as administration, education, mass media, sports and tourism). Oftentimes, this has to do with naming objects pertaining to modern life. This is the case of both the terms that are used in Amazigh for the first time and the ones that display synonymy in the language. In this respect, the question to answer is 'what term to choose?'. Sometimes, recourse is made to borrowings (in Arabic or French) and then an equivalent for the term in question is proposed in Amazigh. As can be noticed, the expansion of the domains of use for Amazigh has been accompanied with a concomitant expansion of terminological demands. One can wonder, in this respect, how CAL can face the challenges of responding to the constantly increasing demands. In other words, how can this plethora of needs be managed? Answering a question like this one will help wonder whether a terminological policy has been placed to respond to new situations of the Amazigh language in the country.

In what follows, the specificity of the corpora used at CAL will be made explicit.

The work done at CAL consists in collecting corpus data with the aim of creating a digitized bank that includes different kinds of data: oral, textual, lexicographical, and terminological.

Indeed, the terminology of an emergent language necessitates a survey of 'what exists' in the first place - *i.e.*, lexical data in use. For the Amazigh language, there is a lexicographical capital which consists of dictionaries, lexicons, and glossaries covering different domains. This lexicographical heritage is marked by the following facts:

- Dictionaries, each covering a particular regional variety.
- The dictionaries containing lexical entries in Amazigh, bilingual on the whole, follow a classification system that differs from author to author: while some dictionaries classify these data by root, others do so by lexemes.
- The graphic system used for the transcription of Amazigh is, depending on the Berberist lexicographer, either Arabic or Latin or, yet, the Tifinagh (De Foucauld).
- The resultant alphabetic classification differs from one dictionary to another. This is because some of the sounds/ graphemes specific to the Amazigh language had to be integrated based either on Arabic or Latin alphabets.
- The development of lexical entries are either equivalents in another language (generally in Arabic or French) in the form of an unelaborated list ("Mirror" presentation or to an Amazigh term that has correspondents in Arabic or French) or a morphological development followed by equivalent(s) in the other language, sometimes with contexts illustrated by proverbs, poetic verses....
- The data takes the attested terms and neologisms, with

no etymology or regional distribution.

In the light of this sketchy overview of Amazigh lexicography, one is called upon to manage variation at different levels, in the absence of a standard code.

1. Management of different orthographies (Arabic, Latin, Tifinagh)
2. Different transcription systems for the same graphic system (IPA, IPA adapted by Semitists, segmentation, phonetic and phonological transcription, different theoretical assumptions (the status of schwa, emphasis))
3. Morphological, syntactic, lexical and regional variation
4. Diversified corpora

The data is necessarily drawn from the terminologist's knowledge, detailed analysis of data, and consideration of different corpora (literary, journalistic, academic, lexical (attested or mixed terms (*i.e.* - attested terms + neologisms))).

Not only management of priorities, but also management of a corpus diversified at different levels.

Mindful of the fact that a bank of linguistic data is both urgent and necessary for fruitful work on terminology, the project in question necessitates as well the reforging or homogenization of data transcription for reasoned exploitation. This course of action requires language mastery such that it can be carried out only by the linguist who knows a variety of dialects, different spelling patterns, different transcription modes (mastery of different theoretical assumptions).

Homogenization of graphic systems may partly be effected automatically, but there will still be a need for another reading in order to correct some elements manually. For example:

1. Schwa epenthesis is consistently present in certain corpora
2. The conversion of the sequence 'gh', which may either correspond to γ or a succession of g+h
3. The application of one segmentation method (orthographic rules).

The extraction of terms issuing from corpora is effected manually: at CAL we still work on paper, in an artisan fashion. This method has the advantage of considering all existing publications to date. Although we might think that this way of doing things is least fruitful at the quantitative level, at the qualitative level, the work is satisfactory on the whole.

Thus, up to now, since the creation of the RSU's at CAL in Mai 2003, the team has managed to come up with 2000 terms, with a mean of 400 terms per annum (Ameur and *al.*, 2006b). We admit that there exists no Amazigh terminological data base, in the proper sense of the term. The data collected are inventoried in an Excel file, which remains the most privileged tool, in the sense that no special knowledge of informatics is required which permits an immediate exploitation in satisfaction of precise, urgent and limited needs. In this respect, the terminological fields relating to the strategic attempt to insert the Amazigh culture and language in the school and media domains. The terms that CAL has been concerned most were the ones relating to the domain of education or the frequently used vocabulary in mass media (the radio and television). The entries studied are processed on a one-by-one basis and the decisions are taken on the basis of a consensus within CAL. In fact, each entry takes around two hours of team work (once all the information is collected and typed on Excel). Then the data is typed on Word, in order to permit subsequent modifications. The data are converted to PDF when the entries are published online on IRCAM's official website or on a "paper" format in pamphlets containing bilingual Amazigh-French or Amazigh-Arabic minimal morpho-syntactic information). The PDF file established, then, is not accessed by the public or other linguists. The targeted audience here are the amazighophones who master at least one variety of Amazigh and who are directly involved in implementing education and mass media (teachers and journalists). The vocabulary and lexical items published up to now are in fact responses to immediate needs of the users/consumers of the language. These publications are thus incomplete in the sense that they target a very specified audience.

Added to this is the fact that the lexicographical data collected by Berberists remains incomplete on the grounds that no dictionary, and no lexicon, has by any means been exhaustive. The lexicon being an open, perpetually changing list, some words die and others are created every day.

7 Conclusion

The revitalization of the Amazigh language in Morocco has enjoyed a positive situation since the creation of IRCAM, which contributed to, among other things, the institutionalization of the language, its implementation in the country's educational system, and the expansion of its domains of use. IRCAM is furthermore conceived of as a legitimate station as far as standardization is concerned, in the sense that it is constantly called upon by different public and private spheres.

The foregoing factors do contribute to the promotion of the Amazigh culture, in particular the Amazigh language.

However, what has been acquired should not blink the fact that much remains to be done in order to preserve and promote the Amazigh language, especially in what concerns the creation of all kinds of digitized data bases. The latter are fundamental tools for a qualitative and quantitative leap for the mission assigned to IRCAM and, in particular, CAL. The linguistic challenges, we have seen, are numerous and various, but are not at all difficult to surmount. Collaboration among researchers of various disciplines is, therefore, necessary in order to benefit from their experience and expertise, especially in the area of terminology.

Acknowledgements

I would like to thank Nourddine Amrous for having translated this contribution in English.

References

- Ameur, M. and *al.* (2006a), *Graphie et orthographe de l'amazighe*, Rabat : Publications de l'Institut Royal de la Culture Amazighe, Série : Etudes n°6.
- Ameur, M. and *al.* (2006b), *Vocabulaire de la langue amazighe (Français-Amazighe) 1*, Rabat : Publications de l'Institut Royal de la Culture Amazighe, Série : Usuels n°1.