

LEXICOGRAPHIE ET COMPARAISON
Le "Dictionnaire informatisé de la langue berbère"

Salem CHAKER (*)

OBJECTIFS ET CONDITIONS GENERALES

Malgré le renouvellement et les développements sensibles de ces dernières années -surtout depuis 1982 avec les publications que nous avons assurées à la SELAF-, le lexique est encore à l'heure actuelle le secteur de la langue berbère le moins bien connu. Il n'existe pas d'instrument lexicographique global, couvrant l'ensemble du domaine et l'on ne dispose même pas d'un dictionnaire des principales racines de la langue. De plus, la couverture lexicale des dialectes reste très inégale, tant au plan quantitatif que qualitatif. Pour nombre de régions importantes -notamment au Maroc-, on ne dispose toujours pas d'un bon dictionnaire. et beaucoup des instruments existants sont vieillis, devenus quasiment introuvables et parfois peu fiables.

(*) LAPMO (CNRS), Aix-en-Provence / I.N.A.L.C.O., Paris.

Cet état de chose est évidemment très préjudiciable à la recherche linguistique berbère, en matière de description synchronique, comme de reconstruction historique : toute la linguistique historique berbère et la comparaison berbère/chamito-sémitique bute actuellement sur l'absence d'un dictionnaire des racines berbères qui permettrait une mise en rapport tant soit peu systématique des lexiques et, subséquentement, l'établissement d'un réseau de correspondances phonétiques ayant une validité générale.

Mais cette lacune se ressent aussi du point de vue des applications : planification linguistique, élaboration d'un "berbère fondamental", d'instruments didactiques..., préoccupations qui sont de plus en plus à l'ordre du jour parmi les populations berbérophones. Les expériences néologiques actuelles -particulièrement vigoureuses en milieu kabyle- pèchent toutes par leur caractère aléatoire et approximatif (emprunt à d'autres dialectes, création à partir de racines locales...). Bien évidemment, ces "ratés", ces bizarreries de la néologie kabyle s'expliquent en grande partie -bien qu'il y ait d'autres causes de nature idéologique- par l'absence de bons outils lexicographiques, accessibles et représentatifs, au moins pour les principaux dialectes berbères.

Il est par ailleurs clair que les conditions actuelles de l'exercice de la recherche ne permettent plus à un chercheur isolé, avec des moyens classiques, de mener à bien une tâche aussi lourde que la réalisation d'un dictionnaire de langue couvrant un ensemble vaste et différencié comme l'est le domaine berbère. Une carrière entière n'y suffirait pas : les temps ne sont plus à des Charles de Foucauld (ou des Littré, Larousse...) consacrant leur vie à l'élaboration d'un dictionnaire d'une minutie et d'une ampleur admirable.

D'autre part, les supports classiques (fichiers manuels) -en admettant que l'on parvienne au terme du travail- ont le grave inconvénient d'être figés, d'une utilisation sans souplesse et de laisser entier le problème de l'édition finale du dictionnaire, dont la publication imprimée reste une opération hautement problématique, d'un coût financier prohibitif et d'une réalisation technique extrêmement délicate. Ces obstacles matériels étant encore plus grands dans un secteur comme le nôtre où le public potentiel n'est pas très large. C'est ainsi que l'on peut comprendre l'existence de plusieurs fichiers de dictionnaires inédits -les plus importants étant ceux d'A. Roux (parlers du Maroc, déposés au LAPMO à Aix) et celui initié par A. Basset (INALCO, Paris). Concrètement, sous leur forme première, de tels fichiers (celui de Roux doit atteindre les 75000 fiches !) sont à peu près inutilisables parce que trop vastes pour être exploités par des moyens traditionnels.

Il s'est trouvé que mon laboratoire de rattachement (LAPMO) disposait d'une documentation berbère exception-nelle, pour partie originale : presque toutes les sources publiées en matière de lexicographie berbère + fonds inédit A. Roux. Et, depuis quelques années, j'ai eu la possibilité d'utiliser les moyens informatiques de l'IREMAM (Institut auquel appartient le LAPMO) pour la réalisation de la bibliographie analytique des études berbères que j'assure régulièrement pour l'Annuaire de l'Afrique du Nord. J'ai pu ainsi informatiser entièrement cette chronique et créer une base de données "Berbère" qui compte actuellement plus de 2000 documents analysés. A travers cette expérience, j'ai ainsi eu l'occasion de découvrir les potentialités de l'informatique documentaire, non seulement en tant qu'outil de gestion de données nombreuses et complexes, mais aussi en tant que puissant instrument de recherche et d'aide à l'édition.

C'est de cet ensemble de constats, d'opportunités et d'expériences qu'est né en 1986-87 le projet de "Dictionnaire informatisé de la langue berbère". Aix présentait des conditions favorables pour concevoir et porter un tel projet de longue haleine. D'autant que mes activités scientifiques depuis 1981 m'avaient permis d'y mettre en place un réseau de collaborations scientifiques assez large : avec certains berbérissants maghrébins (Algérie/Maroc) et avec mes étudiants (notamment ceux de doctorat).

En 1986-87, un crédit spécial du Ministère de la Recherche nous a permis de lancer ce programme, dont les objectifs concrets ont été définis comme suit :

Constitution d'une base de données lexicales berbères significative (contenant le maximum possible de l'information lexicale publiée et des fonds inédits accessibles), qui pourrait atteindre 30 à 50.000 lexies. Cette base lexicale, en extension indéfinie, devant permettre l'interrogation à partir de tous les paramètres d'analyse des documents et constituer un instrument de recherche lexicologique permanent.

Préparation de l'édition de différents dictionnaires berbères : dictionnaires dialectaux, dictionnaire des racines berbères, glossaires thématiques et, à terme, "Dictionnaire de la langue berbère"...

AVANCEMENT ACTUEL DU PROGRAMME

Le programme a désormais atteint sa vitesse de croisière et les trois premières années de fonctionnement ont notamment permis :

– la mise en place effective du réseau de collaborateurs, avec répartition des sources à dépouiller entre les différentes équipes et correspondants.

– l'élaboration d'un bordereau de dépouillement et l'adoption des normes communes d'analyse des matériaux et de remplissage des bordereaux. L'ensemble des problèmes linguistiques ont été longuement examinés et ont fait l'objet de décisions et recommandations précises : notation, traduction française, identification syntaxique et morphologique, identification des variantes, extraction de la racine, illustration des usages par des énoncés, identification des emprunts, indexation thématique des lexies...

– le lancement simultané des opérations de dépouillement à Aix, Fès, Oujda et Tizi-Ouzou. Actuellement, plus de 15000 bordereaux ont été remplis (Aix ± 4500 ; Tizi-Ouzou ± 10000 ; Fès ± 1000 ; Oujda ± 1000).

– l'expérimentation et l'adaptation du logiciel de base de données "Texto" de CHEM-DATA utilisé par l'IREMAM à la nature des matériaux à traiter. L'intérêt fondamental de ce logiciel (outre qu'il nous était déjà familier puisqu'il supporte toutes les bases de données de l'IREMAM) réside dans sa très grande souplesse et sa puissance : il admet jusqu'à 99 champs distincts par document ; de nouvelles zones peuvent être créées à tout moment (la structure du document peut être éventuellement modifiée en cours de développement) ; il permet toutes les manipulations et interrogations simples ou croisées à partir des trois opérateurs logiques (ET/OU/SAUF) et l'indexation sur tous les champs constitutifs du document. Ces possibilités sont essentielles pour une base lexicale puisqu'elles forment un puissant outil de recherche lexicologique et autorisent, à partir d'un seul fichier, la fabrication d'un grand nombre de produits de sortie différents : dictionnaire inverse français/berbère, dictionnaire des racines, dictionnaires dialectaux, dictionnaires thématiques, index morphologiques, index syntaxiques...

– le démarrage des opérations de saisie des matériaux lexicaux (février 1987). La saisie étant assurée grâce à un financement régulier de vacances par l'IREMAM. Pour "cette mise en machine", priorité a été accordée aux matériaux lexicaux provenant du fonds inédit A. Roux (LAPMO) concernant le Maroc Central : nous disposons là d'une source originale pour un dialecte berbère qui est l'un des moins bien connus en matière lexicale. De plus, plusieurs de mes étudiants-collaborateurs étant originaires de cette région, ce choix s'imposait aussi pour des raisons d'efficacité.

Actuellement, les opérations de saisie (suivies de relectures et corrections en machine) se poursuivent à un rythme satisfaisant. A la fin décembre 1988, la base comportait deux fichiers dialectaux : "Maroc central" comptant 3751 entrées lexicales et "Chleuh" ± 500 entrées. Nous préparons actuellement une pré-publication du fichier "Maroc central" que nous espérons pouvoir mettre en circulation en juin ou septembre 1989.

QUELQUES CONSTATS SUR LA STRUCTURE DU LEXIQUE : de l'élaboration du dictionnaire berbère.

Le fichier "Maroc central" qui a donc atteint maintenant une certaine ampleur (3751 entrées) nous permet de formuler plusieurs conclusions quantitatives et qualitatives sur la structure du lexique dans un dialecte berbère nord. et, subséquemment, d'étayer sur des bases numériques, nos options de départ pour l'élaboration du dictionnaire berbère. Ces choix initiaux découlaient des constats auxquels j'avais pu aboutir, il y a quelques années, à partir de mon étude de la dérivation verbale kabyle (1973), ainsi que d'une recherche sur la formation du lexique kabyle menée à bien par l'un de mes étudiants de doctorat en 1985 (Haddadou). Ces conclusions sont d'ailleurs en large recoupement -même si les solutions concrètes retenues divergent partiellement- avec la réflexion récente de M. Taïfi sur ce sujet (1988).

a) L'ensemble des 3751 entrées lexicales du fichier "Maroc central" correspond à 2193 racines distinctes. Ce qui fixe le taux de productivité globale des racines à 1,7 -niveau particulièrement faible pour une langue réputée reposer sur un système de dérivation par racine + schème. Concrètement, cela signifie que chaque radical consonantique produit moins de deux dérivés lexicaux, verbaux ou nominaux.

b) Sur les 3751 entrées du fichier, 1405 seulement ont pu être analysées comme étant des dérivés réguliers, liés par une relation formelle et sémantiques prévisible à leur racine ; soit un taux de 37,5 %. Ce pourcentage nous donne le taux de motivation globale du lexique dans ce dialecte. En conséquence, la très nette

majorité des unités lexicales (62,5%) repertoriées sont des lexèmes isolés, n'entrant pas dans une famille dérivationnelle vivante.

c). Les 3751 entrées se répartissent en 2371 noms (N) et 1203 verbes (V), le reste étant constitué par des adverbes et unités diverses.

Parmi les 2371 noms, 803 seulement ont pu être analysés comme des dérivés réguliers : le taux de dérivation global dans le sous-ensemble des Noms est donc de 33,85 %. Ainsi, plus de 66 % des Noms sont des lexèmes isolés. Dans l'ensemble de ces nominaux dérivés, la répartition des formes est la suivante :

- (209) 26 % sont des Noms d'action abstraits
- (35) 4,5% " Noms concrets
- (177) 22% " Noms d'agent
- (117) 14,5 % " Noms d'instruments
- (171) 21 % " adjectifs
- (94) 12 % " composés

Parmi les 1203 formes verbales, 602 sont des verbes dérivés, soit un taux de dérivation de 50 %, sensiblement supérieur à celui des noms. La répartition des formes est, pour les verbes, encore plus déséquilibrée :

- (244) 40,5 % sont des dérivés en S-
- (74) 12 % " TTW-
- (121) 20 % " M-
- (117) 19,5 % " N-
- (46) 7,5 % " dérivés complexes

Rappelons maintenant, pour permettre une comparaison partielle, les principaux résultats de notre exploration de la dérivation verbale kabyle (Chaker 1973) :

les 1084 verbes dérivés recensés en kabyle étaient issus de 660 racines distinctes, soit un taux global de productivité dérivationnelle de 1,6, un peu plus faible que celui de Maroc central (1,7). Ces 1084 verbes dérivés se répartissaient comme suit :

- (488) 45 % de dérivés en S-
- (190) 17,5 % " en TTW-
- (77) 7% " en M-
- (50) 4,5 % " en N-
- (80) 7,5 % " en MY-
- (199) 28,5 % " complexes

Il y a donc bien quelques divergences quantitatives entre le Maroc central et la Kabylie en matière de dérivation (verbale), mais elles sont nettement marginales et l'on peut affirmer que les grandes tendances sont les mêmes. En un niveau général, on peut conclure, pour les deux dialectes même si les données numériques sont incomplètes pour le kabyle, que :

1. Une très nette majorité du lexique de ces dialectes est constituée par des formes isolées, n'appartenant à aucun champ dérivationnel productif en synchronie.

2. Cette proportion de lexèmes isolés est beaucoup plus forte pour les nominaux que pour les verbes. La dérivation résiste bien mieux dans la sphère verbale.

3. La productivité dérivationnelle moyenne des racines est très faible dans les deux dialectes, nettement inférieure à 2 unités.

Le lexique de ces dialectes est donc, en synchronie, majoritairement non-motivé. C'est-à-dire que la langue est désormais plus "lexicologique" que "grammaticale" : le modèle théorique chamito-sémitique de formation du lexique par combinaison de racines et de schèmes réguliers a perdu l'essentiel de sa validité en berbère nord. La dérivation n'y concerne plus qu'une (grosse) minorité du lexique. En d'autres termes, la notion de racine en berbère est de moins en moins une réalité synchronique, et de plus en plus une donnée diachronique.

Les implications de ces constats au niveau lexicographique sont évidentes et confortent a posteriori les options que nous avons prises pour l'élaboration du dictionnaire berbère :

l'organisation traditionnelle du dictionnaire sur la base des racines n'est plus vraiment fonctionnelle en berbère nord. Il paraît préférable d'y adopter l'ordre alphabétique des unités lexicales réelles, chacune étant

traitée comme entrée autonome. La racine continuera cependant à être prise en compte, en tant qu'information (le plus souvent diachronique) associée à l'unité lexicale effective.

On pourrait éventuellement, comme le suggère Taïfi (1988), établir une distinction entre Nom et Verbe, et maintenir pour ce dernier l'organisation par racine puisque la dérivation y est encore assez productive et qu'elle peut donc servir de base de travail dans le sous-ensemble verbal. Mais on introduirait alors une hétérogénéité dans le principe d'organisation du dictionnaire berbère, hétérogénéité évidemment peu satisfaisante pour l'esprit et dont on peut craindre qu'elle ne complique la consultation et l'utilisation du dictionnaire.

La solution la plus adéquate pour le berbère nord semble donc bien être le classement alphabétique pur, accompagné d'un index systématique des racines.

Bibliographie :

- CHAKER (Salem) : 1973 - Le système dérivationnel verbal berbère (kabyle), Thèse de 3ème cycle, Univ. de Paris-V.
- CHAKER (Salem) : 1984 - Textes en linguistique berbère. Introduction au domaine berbère, Paris, CNRS.
- HADDADOU (Mohand-Akli) : 1985 - Structures lexicales et signification en berbère (kabyle), Thèse de 3ème cycle, Univ. de Provence.
- TAIFI (Miloud) : 1988 - Problèmes méthodologiques relatifs à la confection d'un dictionnaire du Tamazight, AWAL : Cahiers d'études berbères, 4 : 15-26.