

LE DICTIONNAIRE ET L'INFORMATIQUE(*)

N. RICHERT & Dr. G.F. ROMERIO
Expert UNESCO

SOMMAIRE

1. INTRODUCTION
2. L'ELABORATION DES DICTIONNAIRES ET L'INFORMATIQUE
 - 2.1. Généralités
 - 2.2. Constitution d'un réservoir d'informations
 - 2.3. Le travail lexicographique et le temps
 - 2.4. Mise à jour
 - 2.5. Edition
 - 2.6. L'informatique au service des dictionnaires
 - 2.6.1. Constitution d'une base de données lexicographiques
 - 2.6.2. Flexibilité et rapidité des manipulations
 - 2.6.3. Mise à jour
 - 2.6.4. Apport de l'informatique rédactionnelle
 - 2.6.5. Exploitation d'une base de données lexicographiques
3. LES PROBLEMES DES DICTIONNAIRES ARABES
 - 3.1. Les dictionnaires bilingues comme instruments de traduction
- 3.2. Les dictionnaires, source d'information linguistique
- 3.3. Les problèmes des terminologies en langue arabe
- 3.4. Conditions d'une lexicographie arabe moderne efficace
- 3.5. Rôle de l'informatique
4. LE PROGRAMME LEXAR
 - 4.1. Les infrastructures technologiques
 - 4.2. La structuration des données
 - 4.3. La saisie des données
 - 4.4. L'exploitation des données
 - 4.4.1. Pour l'observation et la recherche
 - 4.4.2. Pour la publication lexicographique bilingue
 - 4.4.3. Pour la consultation publique directe
5. CONCLUSION

(*) Séminaire régionale UNESCO de formation professionnelle en matière de gestion de l'édition Rabat 11-20 Décembre 1978/Institut d'Etudes et de Recherches pour l'Arabisation.

1. INTRODUCTION

Il peut être de quelque intérêt dans le cadre de ces journées de réflexion sur l'Édition d'arrêter l'attention des participants sur l'un des produits les plus élaborés et les plus prestigieux de l'industrie du livre : le dictionnaire.

Parmi les productions littéraires des grandes langues de civilisation, s'il y a des œuvres représentatives de la culture et de la science, ce sont bien les dictionnaires en dépit de la place modeste que leur accorde les histoires des littératures.

Qu'ils aient pour but de décrire la langue ou de traduire d'une langue à une autre, qu'ils visent à informer sur l'histoire, la géographie, les arts, les sciences et les techniques, les dictionnaires sont les lieux de référence privilégiés qui rassemblent les patrimoines linguistiques, culturels, scientifiques communs. Ce sont les dépositaires du savoir et par là même des outils indispensables dans la transmission et l'acquisition de l'information au sens large du terme.

A ce titre, ils sont -ou devraient être- l'objet de tous les soins de ceux qui s'occupent de la diffusion du savoir par le livre. Depuis, la conception initiale de l'ouvrage, qui fait appel aux ressources culturelles les plus représentatives, jusqu'à la réalisation pratique qui requiert les compétences techniques les plus raffinées de l'industrie du livre, l'élaboration d'un dictionnaire est une aventure intellectuelle et technique dont le maître d'œuvre et l'éditeur conduisent les destinées.

A notre époque où l'information est devenue aussi précieuse que l'énergie, car d'elle aussi dépend le développement économique, social et culturel des peuples, les dictionnaires jouent un rôle capital car, en attendant que la planète soit quadrillée de réseaux informatisés d'accès à des bases de données mises à la portée de tous, on se servira encore longtemps du livre pour s'informer et, en particulier de ces livres spécialement destinés à enregistrer le savoir que sont les dictionnaires.

Or le développement des sciences et des techniques, l'évolution des langues qui en résulte, sont si rapides, les besoins d'information précise, exacte, complète et à jour sont si grands, qu'il n'est pas possible de concevoir l'élaboration des dictionnaires selon les méthodes artisanales traditionnelles de la lexicographie. Les technologies de stockage et de traitement de l'information peuvent et même doivent être mises au service de ce secteur particulier de l'activité littéraire. C'est là un premier aspect que nous voudrions examiner ici.

Nous aimerions aller au delà en appliquant notre réflexion à la lexicographie arabe contemporaine que nous examinerons du point de vue de l'information qu'elle produit. Par delà un bref examen critique des dictionnaires arabes, nous serons conduits à observer la situation linguistique qu'ils révèlent, situation qui, dans certains domaines, nécessite une révision des méthodes lexicographiques et l'utilisation de la technologie des ordinateurs.

A titre d'illustration, nous donnerons un bref aperçu sur le programme LEXAR dont l'Institut d'Études et de Recherches pour l'Arabisation s'est vu confier la responsabilité à l'échelon du monde arabe.

2. L'ÉLABORATION DES DICTIONNAIRES ET L'INFORMATIQUE

2.1. Généralités

Avant d'étudier plus particulièrement les problèmes d'élaboration des dictionnaires dans la perspective d'utilisation des moyens informatiques, il n'est peut-être pas inutile de rappeler les caractéristiques générales de ces ouvrages, lesquelles nous permettront de situer plus précisément notre réflexion.

Point de vue de l'utilisateur

Rappelons que, pour l'utilisateur, le dictionnaire est un genre particulier de livre, une référence qu'il consulte pour y trouver des informations ponctuelles. Selon la nature de l'ouvrage, le lecteur y cherchera :

- des informations sur sa propre langue (dictionnaires monolingues ou linguistiques) afin de la mieux maîtriser et de mieux communiquer à l'intérieur de sa communauté dont la langue est l'institution;

- des informations sur d'autres langues (dictionnaires et lexiques bilingues ou multilingues) afin de communiquer avec des communautés extérieures à la sienne, de pouvoir comprendre et/ou traduire;

- des informations sur le monde (encyclopédies ou dictionnaires encyclopédiques, dictionnaires de spécialités) afin de se documenter ou de parfaire ses connaissances dans des domaines divers.

Quel qu'il soit, le dictionnaire fonctionne comme une **mémoire** : il donne des réponses à des questions. Les réponses qu'il donne ont pour le lecteur une valeur **contragnante** : ce que dit le dictionnaire est non seulement **exact** mais aussi **obligatoire**. Car que ce soit dans le domaine du

savoir sur la langue ou celui du savoir sur le monde, le dictionnaire est le réceptacle du patrimoine commun. A ce titre il est en institution sociale qui doit garantir l'intégralité et la qualité de l'information transmise aux membres de la communauté.

Point de vue de producteur

Nous n'insisterons pas sur le fait qu'en tant que produit de l'industrie du livre, le dictionnaire obéit aux règles générales de ce type de production : marché éventuel de lecteurs dont on étudie les besoins et le pouvoir d'achat, contraintes commerciales qui commandent les dimensions de l'ouvrage, etc. Sans négliger ces facteurs socio-économiques, nous nous attarderons de préférence à un aspect qui nous paraît fondamental et original dans la réalisation des dictionnaires : les producteurs, auteur et éditeur, créent d'abord un **instrument d'accès à l'information**; la nature de cette information, sa quantité, sa qualité et son coût orientent dès le départ l'avenir du produit, non pas seulement son avenir commercial mais aussi son avenir culturel, en tant qu'objet institutionnel.

La fabrication de cet instrument d'accès à l'information qu'est le dictionnaire nécessite des opérations spéciales et entraîne des contraintes particulières dont nous rappellerons brièvement la nature.

2.2. Constitution d'un réservoir d'informations

Quelle que soit la dimension de l'ouvrage ou sa destination, un dictionnaire rassemble des informations discontinues rangées dans un ordre précis. en général l'ordre alphabétique, convention arbitraire mais commode pour classer et retrouver l'information.

Ces informations sont toujours nombreuses, de l'ordre de plusieurs milliers d'unités.

Par exemple, le dictionnaire «petit Robert» de la langue française compte 47.000 entrées, le Petit Larousse 70.500. Si l'on compte en nombre de caractères typographiques, les données s'élèvent à plusieurs millions : 21 millions de signes pour l'impression du Petit Robert, par exemple.

La nature des informations, qui sont toujours linguistiques (à l'exception des éléments illustratifs) est complexe et structurée. Un article de dictionnaire n'est pas un texte libre : c'est un «programme» au sens technique : une suite ordonnée d'informations formulées dans un langage codé. L'information totale se présente presque toujours sous la forme d'une double structure hiérarchisée :

- une structure principale : la liste des mots ou «entrées» rangées dans l'ordre choisi et qui forme l'armature principale du dictionnaire;
- Une structure secondaire : les informations sur les entrées, dont le contenu peut être variable mais dont le mode de présentation suit un schéma préétabli toujours identique à lui-même.

La qualité des informations n'est évidemment pas indifférente. L'exactitude est, on s'en doute, prépondérante. Dans le dictionnaire, plus que partout ailleurs, l'erreur est ressentie comme une faute grave. Il est donc nécessaire d'exercer un contrôle rigoureux de l'information, ce qui ne peut se faire sans la mise en place d'un appareil documentaire important (références bibliographiques ou autres, citations, etc.), qui, même s'il n'apparaît pas dans la publication finale est indispensable à l'élaboration. Il faut aussi à tout moment vérifier l'information ce qui nécessite la mise en place de procédures de révision et de corrections à tous les stades d'élaboration.

Trois caractéristiques se dégagent de ce premier aspect de l'élaboration des dictionnaires :

- grande quantité et complexité des informations discontinues;
- organisation systématique et structuration des informations;
- contrôle documentaire et vérification.

On conçoit que, par les méthodes artisanales classiques, la collecte et le recensement des matériaux de base d'un dictionnaire fasse appel à un personnel nombreux et très compétent, ou, dans le cas d'entreprises individuelle, à de nombreuses années de travail. Un dictionnaire moderne est cependant toujours une œuvre collective et même dans le cas de très grand ouvrages, une entreprise conçue et menée selon une planification industrielle rigoureuse.

L'exemple du Grand Larousse Encyclopédique de la langue française est à cet égard hautement instructif. Paru en 1960, cet ouvrage contient 163.270 articles qui couvrent plus de 400.000 acceptions de termes. Son élaboration qui a demandé une dizaine d'années de travail et a nécessité la collaboration de près de 2000 personnes a été organisée de la manière suivante :

- 1°) Etablissement d'une classification des connaissances afin de délimiter les vocabulaires des spécialités : 711 rubriques de sciences humaines et 529 rubriques de sciences exactes et fondamentales ont été définies.

2°) Chacun des 1300 vocabulaires spéciaux déterminés par ces rubriques a été confié à un spécialiste de la discipline pour qu'il en dresse la nomenclature à partir d'un dépouillement fait selon des normes fixées à l'avance et pour qu'il procède à une prérédaction.

3°) La révision et la rédaction définitive ont été confiées à un Secrétariat de rédaction comportant 19 Secrétaires qui se partageaient 19 grandes disciplines telles que Langue, Histoire, Géographie, Art, Musique, Cinéma, Sciences juridiques, etc.

4°) Suivirent des procédures de vérification, de normalisation des textes, et de relecture par le rédacteur en chef avant les opérations de composition, elles-mêmes suivies de plusieurs vérifications.

En ne comptant que les opérations principales, on atteint le total de 11 procédures de vérification avant la sortie définitive de l'ouvrage.

Il va sans dire que dès 1956, la maison Larousse a mis sur ordinateur toutes les données de base de ce dictionnaire afin de conserver sous une forme maniable et d'exploiter de manière plus systématique ce gigantesque effort collectif.

2.3. Le travail lexicographique et le temps

Il est évident que la manipulation de grandes quantités de données qui doivent être inlassablement vérifiées demande du temps.

L'Organisation Internationale pour la Normalisation (ISO = International Standardization Organization) qui a publié des normes à l'usage des spécialistes qui veulent élaborer des vocabulaires techniques multilingues, recommande de ne pas dépasser le millier de termes techniques en raison principalement de la lenteur du travail. En effet, compte tenu des délais nécessaires, un ouvrage risque d'être caduc au moment de sa parution, la technique dont il s'occupe ayant pu évoluer entre temps.

Le facteur temps est particulièrement crucial dans les ouvrages bilingues ou multilingues car il faut alors traiter les informations et les présenter dans deux ou plusieurs sens. Un dictionnaire français-anglais, par exemple, doit comporter une structure anglais-français où les termes partent de l'anglais (langue source) vers le français (langue cible) et une seconde structure français-anglais où les termes partent du français vers l'anglais. Les duplications qu'entraîne cette nécessité dévorent du temps. Citons l'exemple du Harrap's New Standard French and English Dictionary, nouvelle édition entièrement révisée, dont la partie

français-anglais a été publiée en 1972 mais dont la partie anglais-français, prévue pour 1978 n'est pas encore sortie.

La lenteur et la difficulté du travail lexicographique bilingue sont peut-être les raisons lesquelles les dictionnaires arabes bilingues généraux ne comportent pour ainsi dire jamais la double face (langue européenne - arabe et arabe-langue européenne) quelle que soit la langue source (arabe ou langue-européenne).

Ces brèves considérations sur le temps nous conduisent à un problème spécifique des dictionnaires : leur mise à jour.

2.4. Mise à jour

Il nous faut ici ouvrir une parenthèse sur un phénomène de notre époque qui aggrave considérablement le problème : l'évolution extrêmement rapide des sciences et des techniques et la véritable explosion terminologique qu'elle entraîne. Les langues technologiques voient se développer à grande vitesse des langages particuliers (les terminologies) qui véhiculent l'information scientifique et technique : environ 700 termes techniques nouveaux apparaissent chaque année. Parmi eux environ 4000 se stabilisent et entrent dans la langue.

Cette prévolution terminologique entraîne une activité de traduction sans précédent, la course à l'information scientifique et technique passant par la traduction des terminologies des nations détentrices de l'avance technologique et scientifique. Or pour pouvoir traduire efficacement et rapidement, il faut des instruments fournissant une information complète et à jour.

Une étude menée à l'Université Carnegie Mellon (USA) sur la publication des dictionnaires français-anglais montre que, même dans un domaine en rapide évolution comme l'électronique, un dictionnaire paraît en moyenne tous les 30 mois. Or il s'agit là de nouvelles parutions et non de mises à jour. Celles-ci sont encore plus lentes pour des raisons faciles à concevoir :

L'investissement initial en compétences et en temps que nécessite la publication d'un dictionnaire ne peut être amorti sur une courte période. Par ailleurs, la rigidité de la structure d'un dictionnaire, classé par ordre alphabétique, ne permet pas d'ajouter de la matière à chaque réédition sans soit refondre complètement l'ouvrage, soit fabriquer un supplément qui est en fait un nouveau dictionnaire annexé au précédent.

2.5. Edition

La nature des dictionnaires qui doivent rassembler une matière complexe, rigoureusement contrôlée, sous une forme compacte et maniable, lisible sans difficultés par tous, en rend la réalisation particulièrement délicate et coûteuse.

Pièce maîtresse d'une maison d'édition, le dictionnaire met à contribution les meilleures ressources graphiques et techniques du service de fabrication depuis la composition jusqu'à l'impression finale. Il n'est guère nécessaire d'insister sur ce point sinon pour rappeler que dans ce domaine aussi le dictionnaire exige du temps, un personnel nombreux et qualifié et de longues et minutieuses mises au point et vérifications.

Dans ce bref panorama des différents aspects de l'élaboration des dictionnaires, nous avons tenté de dégager certaines caractéristiques fondamentales qui sont :

- grande quantité et complexité des informations à rassembler et qui constituent les matériaux de base;
- complexité des opérations qu'exigent leur mise en ordre et leur structuration;
- lenteur des procédures d'exécution à tous les stades;
- exigence fondamentale de rigueur qui nécessite un contrôle permanent de l'information à tous les stades d'élaboration.

2.6. L'informatique au service des dictionnaires

La technologie des ordinateurs, dont les progrès sont extrêmement rapides, permet aujourd'hui de traiter de l'information non numérique et en particulier de stocker, structurer et restituer des messages linguistiques de toutes natures dans des conditions particulièrement intéressantes pour les problèmes qui nous préoccupent.

2.6.1. Constitution d'une base de données lexicographiques

L'informatique permet de rassembler et recenser une quantité énorme de données à condition que les informations introduites aient été préalablement structurées.

Les techniques de stockage et de traitement de données linguistiques ont maintenant été largement expérimentées, notamment dans les banques de données documentaires, dont les principales rassemblent jusqu'à plusieurs dizaines de millions

d'unités documentaires. Celles-ci sont constituées de références bibliographiques, dont de messages linguistiques de longueur et de nature variable.

Nous avons déjà cité l'exemple de la maison Larousse qui dès 1956 a mécanisé les informations entrant dans la constitution du Grand Larousse Encyclopédique.

La nature de l'information lexicographique, structurée et récurrente se prête particulièrement bien à l'utilisation de techniques de stockage et de traitement dérivées des techniques documentaires informatisées. En effet, l'ordinateur permet de stocker de très nombreuses unités discontinues et peut emmagasiner en mémoire l'information plus ou moins complexe accompagnant ces unités, à savoir dans le cas du dictionnaire de langue, par exemple, la programme habituel d'un article de dictionnaire : indications grammaticales, prononciation, étymologie, datation, définition, exemples, etc., ou dans le cas d'un dictionnaire bilingue : les équivalents de l'entrée, les exemples et les idiotismes.

Il est aussi possible sans surcharge excessive de stocker avec les unités linguistiques tout un appareil documentaire indispensable tel que références bibliographiques, sources des unités, contexte, nom du responsable traitant ces unités, date de traitement, dates des vérifications, etc. Cet appareil documentaire nécessaire à tout travail lexicographique sérieux est particulièrement difficile à mettre en place et à manier dans l'élaboration manuelle traditionnelle.

En outre, la souplesse de l'instrument informatique permet de stocker une information provisoirement fragmentaire au fur et à mesure de sa disponibilité et de l'enrichir ou de la compléter ultérieurement, puisque tout complément d'information peut être introduit à sa place à n'importe quel moment et sans recherches manuelles fastidieuses.

2.6.2. Flexibilité et rapidité des manipulations

En effet, une des caractéristiques des procédures d'automatisation est la prise en charge par la machine, à des vitesses et avec une infaillibilité inaccessibles à l'homme, d'opérations longues, fastidieuses et génératrices d'erreurs lorsqu'elles sont faites manuellement.

Tous les classements et les tris peuvent ainsi être faits mécaniquement, non seulement pour les unités principales mais pour tout élément du programme d'informations qui les accompagne. Par exemple, il est possible de demander à l'ordinateur

de dégager dans une nomenclature donnée, tous les noms communs accompagnés de pluriels irréguliers, afin de les vérifier ou pour toute autre opération nécessaire.

Autre exemple concernant une structure bilingue, français-arabe, supposons : l'ordinateur peut inverser les relations et ranger toutes les unités arabes par ordre alphabétique avec tous les équivalents français et donc retourner la structure sous forme arabe-français, quitte bien entendu à faire des réaménagements nécessaires pour assurer la cohérence et la symétrie de l'ensemble.

A l'intérieur d'une même base de données lexicographiques, il est possible de faire des tris sectoriels automatiques : dégager le vocabulaire de l'agriculture ou de la médecine, si au départ, bien entendu les informations sur les domaines d'emploi des termes ont été dûment enregistrées.

Il est donc possible d'intervenir de façon multiforme, ponctuelle ou systématique sur les données enregistrées : recherche, sélection, classement, tri, contrôle de cohérence, etc, dans des conditions optimales de rapidité avec un moindre risque d'erreur, la seule condition relative à ce dernier point étant que les données initiales aient été correctement entrées.

2.6.3. Mise à jour

La flexibilité d'accès d'une base de données informatisées ouvre la voie à une mise à jour facile. Comme il est possible d'intervenir à n'importe quel point de l'ensemble et à n'importe quel moment, la mise à jour ne présente guère de difficultés. Conçue comme un matrice en constante évolution, la base de données lexicographiques peut suivre de façon plus rapide et plus efficace l'évolution linguistique, scientifique et technique.

2.6.4. Apport de l'informatique rédactionnelle

Comme il est techniquement possible d'informatiser la composition typographique des textes, on mesure aisément l'immense progrès qui peut alors s'accomplir dans l'élaboration, la publication et la mise à jour des dictionnaires. La composition mise en mémoire est, elle-même, modifiable dans des délais beaucoup plus courts que ceux des révisions habituelles des dictionnaires. A chaque réédition, il est possible d'introduire une nouvelle information ou de modifier une information existante, ce qui supprime l'inconvénient des suppléments et réduit considérablement le coût des opérations.

2.6.5. Exploitation d'une base de données lexicographiques

Les types d'ouvrages qui peuvent être publiés à partir d'une base de données lexicographiques suffisamment large sont fonction de l'imagination des éditeurs et des besoins du public. Mais comme l'ordinateur peut faire des tris et des sélections programmables extrêmement variés, on peut constituer à partir du stock de données des sous-produits très divers :

- des dictionnaires de spécialités;
- des dictionnaires linguistiques (dictionnaires de prononciation, de rimes, de synonymes, etc).
- des lexiques bilingues ou multilingues (si la base de données comprend une ou plusieurs langues).

Bréf, les possibilités sont d'autant plus grandes que l'investissement initial en information aura été plus large.

3. LES PROBLEMES DES DICTIONNAIRES ARABES

Indépendamment des problèmes d'élaboration déjà évoqués et qui sont communs à tous les dictionnaires, nous aimerions cerner d'un peu plus près la réalité lexicographique arabe en particulier du point de vue de deux grands besoins complémentaires de l'usager : besoins d'information sur la langue arabe contemporaine et besoins terminologiques propres à la traduction.

Il n'est guère nécessaire d'insister sur le fait que la transmission de l'information scientifique et technique et les transferts technologiques qui en découlent se font d'abord par le moyen de la langue. Dans ces domaines l'activité traductrice du monde arabe est essentiellement orientée des langues étrangères vers l'arabe. Cette activité ne peut s'exercer avec profit sans deux types d'instruments de référence : des dictionnaires bilingues qui fournissent les équivalents nécessaires aux termes scientifiques et techniques étrangers, et les dictionnaires monolingues modernes qui renseignent l'usager sur les termes contemporains.

Précisons encore que par traduction nous n'entendons pas ici une transposition toujours possible d'un discours d'une langue dans un discours d'une autre langue, mais le transfert rigoureux et adéquat de l'information, surtout dans les domaines qui exigent rigueur et adéquation, c'est-à-dire les domaines scientifiques et techniques.

est traduit par **العصر الحديث** alors que l'équivalent arabe choisi sert aussi à traduire **période** qui est une subdivision de l'ère dans la terminologie de la géologie.

Il serait faux de croire qu'il s'agit là d'un phénomène exceptionnel, essentiellement dû à un travail lexicographique médiocre. Il est en fait si généralisé dans certains domaines scientifiques et techniques qu'il est nécessaire de contrôler l'information fournie par les dictionnaires monolingues.

3.2. Les dictionnaires arabes, source d'information linguistique

Les dictionnaires monolingues, se divisent en deux grandes catégories :

D'une part, les ouvrages fort anciens qui servent toujours de référence et de norme linguistique, comme, par exemple *Lisân-1-3arabe* (datant du 14ème siècle) ou *Tâj-1-3arous* (du 17ème siècle) où il est parfaitement vain de chercher les sens modernes des termes ou des informations sur le vocabulaire arabe de création récente.

D'autre part, les dictionnaires plus récents, comme *Matn-1-Lugha*, de Ahmed Rida, Beyrouth 1958, *Al Wasît*, de l'Académie du Caire, 1960 ou le célèbre *Al Munjid*, régulièrement réédité depuis 1909. Ces ouvrages surtout généraux, à forte prédominance littéraire, qui décrivent l'usage arabe du 4ème au 20ème siècle sans datation précise, sont fort pauvres en termes scientifiques et techniques contemporains et leur mise à jour souffre du retard habituel aux dictionnaires, aggravé par l'absence de centralisation des faits linguistiques dans le monde arabe.

Ce panorama extrêmement succinct de l'état de la lexicographie arabe contemporaine nous conduit à constater qu'au regard des besoins en information linguistique que nécessite le développement des sciences et des techniques, les instruments d'accès à cette information sont dangereusement insuffisants pour la communauté des 90 millions d'arabophones.

Bien que la lexicographie arabe contemporaine puisse sans nul doute être améliorée, notamment à l'aide des techniques de traitement de l'information, nous ne pouvons passer sous silence un problème grave que révèle l'examen des dictionnaires. En effet, ceux-ci, pour être insuffisants et en retard sur l'évolution de notre temps, n'en sont pas moins les témoins d'un certain état du vocabulaire arabe contemporain face aux langues étrangères technologiques. Dans certains domaines, ce dernier pré-

sente des caractéristiques telles qu'elles nécessitent une approche entièrement nouvelle de l'élaboration des dictionnaires.

3.3. Les problèmes des terminologies en langue arabe

Les observations qui suivent sont un résumé de l'analyse comparative des dictionnaires arabes qui a été conduite par le Professeur Lakhdar Ghazal et dont un compte-rendu est donné dans son ouvrage «*Méthodologie Générale de l'Arabisation de Niveau*», Rabat, 1976 2ème édition.

Sous l'impulsion et la direction du Professeur Lakhdar-Ghazal, l'Institut d'Arabisation a procédé à un dépouillement systématique de toutes les publications bilingues ou trilingues (arabe/anglais/français) parus depuis plus d'un siècle : dictionnaires, lexiques, listes des Académies de langues arabe, publications du Bureau de Coordination de l'Arabisation, etc.

Les études comparatives auxquelles ont procédé le Professeur Lakhdar-Ghazal et les lexicographes de l'Institut ont montré que le vocabulaire arabe contemporain dans les domaines scientifiques et techniques -tel que le révèle l'usage enregistré dans les dictionnaires- est affecté de perturbations qui nuisent considérablement à son pouvoir de transmettre l'information de manière adéquate et à l'égalité avec les grandes langues technologiques.

Nous pouvons les résumer comme suit :

1°) Instabilité du vocabulaire dans la dénomination des notions

Outre les exemples déjà donnés, nous soumettons aux participants un tableau des termes de la classification des êtres vivants (terminologie zoologique et botanique) dans lequel ils pourront aisément constater ce phénomène.

Cette instabilité de la dénomination qui affecte des formes diverses dans de nombreux domaines peut provenir de différents facteurs dont les principaux sont :

- L'absence d'unification linguistique à l'échelon du monde arabe causée par la référence à des langues étrangères différentes (anglais/français notamment) ou par des variations locales de pays à pays ou à l'intérieur d'un même pays;

- L'absence de normalisation effective : malgré les réunions de spécialistes ou les décisions des instances compétentes (Académies de Langue Arabe, Union des Académies ou Organisations

interarabes) les termes scientifiques et techniques se fixent difficilement alors que des termes correspondant à des notions largement diffusées par les mass media se stabilisent rapidement. Il y a là un problème sérieux de diffusion de l'information terminologique dont les auteurs et les éditeurs de dictionnaires devraient tenir compte.

2°) Lacune et vides

Dès que les terminologies atteignent un niveau élevé de technicité, les équivalents arabes sont difficiles à trouver voire introuvables dans les publications courantes et même spécialisées, ces dernières reprenant souvent des termes ordinaires qui ne correspondent pas toujours aux informations fournies par les dictionnaires généraux.

Ces quelques considérations sur les terminologies nous conduisent à réfléchir sur les conditions d'une lexicographie arabe tenant compte des données existantes du vocabulaire arabe contemporain.

3.4. Conditions d'une lexicographie arabe moderne efficace

Il est d'abord évident que pour songer à élaborer des dictionnaires il faudrait disposer au préalable d'un panorama général du vocabulaire arabe contemporain dans tous les domaines et couvrant l'usage de tous les pays arabes.

Le gigantesque travail de recensement d'une langue nationale qu'on entrepris nombre de pays avancés, dont la France, par exemple, qui constitue un Trésor de la Langue Française entièrement informatisé, ne peut être conçu pour l'arabe sans l'aide de l'informatique et, bien entendu, la coopération de toute la communauté arabe.

Mais il faut aussi pouvoir parer au plus pressé et pour cela satisfaire sans retard les besoins en information terminologique créés par la traduction. Avant de décrire le programme LEXAR qui est une première réponse à cette urgence, il nous paraît capital de poser les principes sur lesquels doit reposer tout travail lexicographique bilingue ou multilingue ayant pour base la langue arabe, principes dont il serait souhaitable que les auteurs et les éditeurs de dictionnaires puissent faire leur profit.

1°) Lexicographie critique

Tout projet de dictionnaire ou de lexique devrait partir d'un examen critique préalable des ouvrages existants et des sources attestées du vocabulaire concerné. Car quel que soit le domaine traité, il faudrait :

- dégager les termes sur lesquels il y a accord général interarabe;
- dégager les termes où il y a absence de normalisation;
- détecter les lacunes et les vides.

Etant donné les perturbations des terminologies, un dictionnaire ne devrait pas se concevoir sans un appareil critique de références précises (origine géographique et bibliographique de termes) et d'évaluation du degré d'unification ou de normalisation des termes.

2°) Lexicographie corrective.

Qu'il s'agisse d'usage erroné ou de dialectalismes plus ou moins attestés, il convient de signaler pour les proscrire ou les déconseiller tous les écarts par rapport à la norme linguistique unificatrice qui est l'arabe classique attesté dans la tradition ou créé sur la «qyas»

3°) Lexicographie représentative

L'on ne peut guère proscrire l'individualisme, même dans ce genre littéraire spécial qu'est le dictionnaire. Toutefois, comme ce livre est représentatif d'une communauté entière et non de son seul auteur ou maître d'œuvre, il convient de s'entourer des garanties nécessaires et de faire appel au plus large consensus avant de diffuser les œuvres. Il est en particulier indispensable que les auteurs indiquent précisément leurs inventions. C'est la moindre des choses mais ce n'est hélas pas toujours la règle. Nous ne saurions trop insister, à ce propos, sur la responsabilité des comités de lecture des maisons d'édition, qui, dans l'examen des dictionnaires, pourraient aussi faire appel au patronage d'instances culturelles compétentes.

3.5. Rôle de l'informatique

Si l'on veut vraiment améliorer les techniques d'élaboration des dictionnaires arabes, satisfaire les exigences énoncées ci-dessus et rentabiliser l'entreprise, il est quasi impossible de n'avoir pas recours à la technologie de traitement de l'information dont nous avons préalablement évoqué les principaux atouts.

Sans en réitérer la nature, nous aimerions à titre

d'illustration, exposer les grandes lignes du programme LEXAR, entrepris par l'IERA et qui vise à la constitution d'une base de données lexicographiques multilingues, apparentée aux banques de données linguistiques bilingues ou multilingues mises en place par certains pays (Banque de terminologie de Montréal, Canada; EURODICAUTOM, dictionnaire automatique multilingue de la Commission des Communautés Européennes, etc).

4. LE PROGRAMME LEXAR

Première étape vers un dictionnaire arabe multilingue informatisé, le programme LEXAR au nom acronymique formé à partir de l'expression «LEX ARabes» a pour but la mise sur ordinateur des fichiers bilingues et multilingues de l'IERA que nous avons brièvement évoqués précédemment.

Constitués de fiches de format 10 cm x 15 cm, écrites à la main, ces fichiers contiennent plus d'un million de relations sémantiques entre termes européens (français, anglais, latin scientifique) et termes équivalents arabes, recueillies dans les dictionnaires et les publications lexicographiques d'un monde arabe au cours d'une vingtaine d'années de travail.

Ce stock d'informations lexicographiques, unique en son genre, est difficilement maniable dans sa forme actuelle de quelques 200 tiroirs en bois dans lesquels les fiches sont classées de A à Z. En outre, comme les relations se présentent toutes dans le sens langue européenne-langue arabe, pour pouvoir exploiter parfaitement le fichier, il faudrait pouvoir le renverser intégralement dans le sens arabe-langue européenne, ce qui représenterait un travail manuel gigantesque.

Dans le but d'assurer la conservation de ce réservoir d'informations et de l'exploiter de manière rationnelle pour les travaux de terminologie, l'informatisation des fichiers a été jugée nécessaire à la suite d'une expertise de l'UNESCO. La solution préconisée par l'expert en informatique a été la constitution d'une banque de mots orientée vers l'utilisation lexicographique (élaboration ultérieure de dictionnaires et de lexique). Le programme LEXAR a reçu l'approbation des instances arabes (recommandation de la Conférence CASTARAB, Rabat, 1976, et résolution de l'ALECSO à Khartoum en août 1978) et fait l'objet d'une assistance du Programme des Nations Unies pour le Développement et de l'UNESCO.

4.1. Les infrastructures technologiques du programme LEXAR

Pour pouvoir constituer une banque de mots du type souhaité il fallait d'abord :

1°) Pouvoir utiliser deux alphabets complets pour le stockage des données en caractères latins (y compris les lettres à accent du français) et en caractères arabes (y compris tous les signes de voyellisation).

Le système ASV-CODAR, réductible aux contraintes de l'informatique et compatible avec l'alphabet latin répondait à cette exigence de base. Il est à noter que, pour l'instant, ce système est le seul capable de permettre le traitement et la transmission de données linguistiques arabes entièrement voyellées.

2°) Pouvoir disposer d'organes d'entrée et de sortie capables de traiter les alphabets ainsi définis. Le terminal bi-alphabétique EURAB doté d'une imprimante a été mis au point par l'Institut Européen de Recherche Spatiale (ESRIN) de l'Agence Spatiale Européenne et est disponible pour effectuer les opérations de saisie des données du programme LEXAR.

3°) Pouvoir disposer d'une puissance de calcul et d'un système informatique capable de gérer une banque de données comprenant plusieurs millions de caractères (300 Méga octet). A l'heure actuelle, seul le centre de calcul de l'ESRIN, situé à Frascati (Rome) est capable de répondre à ce besoin, dans les conditions requises par les travaux de l'IERA.

Relié au Maroc par le Centre National de Documentation de Rabat, le centre de calcul vient d'être connecté à l'IERA par le prolongement de la ligne FRASCATI-CND.

4.2. La structuration des données

Bien qu'elles soient toutes de nature lexicographique, donc déjà partiellement mises en forme, les données des fichiers doivent être plus précisément structurées ou selon l'expression technique «formatées» avant d'être introduites de façon acceptable par un calculateur électronique.

Chaque fiche manuelle contient en général un terme en français, son équivalent ou ses équivalents arabes (parfois l'équivalent anglais) et, selon la nature de l'ouvrage dépouillé, des renseignements complémentaires (exemple, définition, etc). Ceci n'est toutefois pas assez structuré pour la saisie.

Aussi un groupe d'analystes est spécialement chargé de transférer les informations contenues dans les fiches manuelles sur des bordereaux de saisie où les informations doivent être inscrites dans des formats préétablis.

Le Bordereau L. 1.2.3.

Le bordereau L.1.2.3. est destiné à consigner une relation sémantique bilingue et comprend deux parties :

- une partie réservée à la langue source (français, en général)
- une partie réservée à la langue cible (arabe)

Ces deux parties sont reliées par un numéro d'accession.

Description du bordereau :

- Une section documentaire (l'étiquette) est commune aux deux parties : elle est réservée, outre au numéro d'accession, aux références documentaires : source de la relation (tel ou tel dictionnaire ou lexique), et au nom de l'analyste ainsi qu'à la date de l'analyse.

1^{re} partie du bordereau : Elle comprend :

- une section réservée au terme de la langue source : l'entrée (langue source).
- une section réservée aux renseignements sur ce terme : l'Indicatif renseignements lexicographiques habituels tels que catégorie grammaticale, genre, nombre, usage, niveau de langue, morphologie.
- une section intitulée **commentaire**, réservée à toute information complémentaire sur l'entrée : définition, exemple, domaine d'emploi, etc.

2.2. Constitution d'un réservoir d'informations

- d'une première ligne réservée à une appréciation critique de la source ou à une norme (terme unifié, terme recommandé par tel ou tel congrès, etc)

et se compose de

- une section réservée à l'équivalent arabe de l'entrée langue source : l'entrée (langue cible)
- une section pour les renseignements sur ce terme : **indicatif** (langue cible).
- une section réservée au **commentaire** (comme dans la première partie).

On aura noté que le bordereau est conçu de façon à respecter l'équivalence des langues dans un sens ou dans un autre, sens français-arabe, ou arabe-français.

Les langues supplémentaires, l'anglais et le latin scientifique, sont enregistrées dans un bordereau identique : L.1.2.3. mais monolingue comprenant les trois sections déjà décrites (Entrée, Indicatif et Commentaire). La relation trilingue ou quadrlingue est restituée par le numéro d'accession

identique pour tous les équivalents d'une même relation.

4.3. Saisie des données

Une fois les bordereaux remplis, leur contenu est transféré en mémoire par l'intermédiaire des terminaux d'ordinateur, lesquels permettent de visualiser les données introduites et d'opérer les corrections nécessaires et de donner des instructions à la mémoire centrale pour le stockage et/ou la restitution des données.

4.4 L'exploitation des données

4.4.1. Pour l'observation et la recherche

Une fois les données mises en mémoire, il sera possible de les exploiter de diverses manières. Indépendamment de tous les tris et classements que l'on peut imaginer, les opérations majeures que permettra la mécanisation peuvent être résumées comme suit :

1°) Renverser toutes les relations selon la langue choisie entre toutes les langues du fichier et principalement de l'arabe vers les langues européennes, opération de retour indispensable pour détecter les perturbations de la terminologie en langue arabe.

2°) Dégager à partir de reconnaissances statistiques de fréquence des termes arabes, la partie non perturbée des terminologies, là où il y a unanimité des traductions.

3°) Dresser les faisceaux de termes perturbés dans des champs lexicaux donnés et diagnostiquer la nature des perturbations. Précisons ici qu'ont été conçus deux bordereaux supplémentaires d'analyse qui prévoient une étude plus approfondie de certains termes scientifiques et techniques, laquelle a pour but de fournir aux linguistes des informations complémentaires permettant d'étudier la normalisation de termes perturbés.

4°) Dégager les zones lexicales où la terminologie arabe présente des insuffisances de traduction, termes étrangers expliqués et non traduits, notions différentes confondues sous une même appellation, lacunes totales.

4.4.2. Pour la publication lexicographique

Deux types de publications peuvent être prévues :

1°) Publications de recherche et de consultation

Les perturbations constatées et étudiées par l'IERA pourront faire l'objet de documents de tra-

vail soumis au consensus arabe par l'intermédiaire des organisations compétentes (ALECSO, Bureau de Coordination de l'Arabisation, Académies de langue arabe) pour étude, appréciation et complément d'information.

Les renseignements recueillis pourront être centralisés et largement diffusés afin de faciliter les procédures de normalisation des termes techniques et scientifiques arabes.

2°) Publications généralisées

Le vocabulaire bilingue ou trilingue unifié qui aura été dégagé par le programme LEXAR pourra donner lieu à une publication soit globale soit sous forme de lexique sectoriels, par domaines d'emploi.

Tous les travaux terminologiques qui auront fait l'objet d'un large consensus pourront eux aussi être largement diffusés.

4.4.3. Pour la consultation publique directe

Indépendamment de tous les chercheurs ou des organismes qui pourraient en faire la demande, la banque LEXAR qui aura été élaborée à partir de la base de données lexicographiques que nous avons décrite précédemment sera à la disposition de tout le public arabe comme aide à la traduction au même titre que les banques similaires existant dans le monde.

5. CONCLUSION

Il est à remarquer que le programme LEXAR, spécialement orienté vers la lexicographie bilingue ou multilingue est conçu dans l'optique de permettre une remise en ordre et une mise à jour des

terminologies arabes sur la base d'une confrontation avec les langues européennes par le biais de la traduction.

Il s'inscrit par là dans le mouvement général de l'activité terminologique internationale. Chaque pays que préoccupe l'avenir technologique de sa langue, c'est-à-dire la conservation du patrimoine culturel sans sacrifice de la modernité, exerce aujourd'hui cette activité qui consiste à mettre au point les équivalents en langue nationale des concepts nouveaux nés du progrès scientifique et technique. La nécessité de traduire si l'on veut suivre l'évolution du monde contemporain conduit donc à exercer une action sur la langue afin d'améliorer sa capacité à véhiculer efficacement le savoir.

Ces activités linguistiques ne se conçoivent plus sans l'utilisation de l'informatique seule capable de répondre aux besoins quantitatifs et qualitatifs avec la rapidité et la flexibilité nécessaires. Terminologie, lexicographie et informatique sont désormais liées pour une meilleure transmission de l'information. Cela est si vrai que nous voudrions rappeler en conclusion que le Centre International d'Information sur la Terminologie (INFOTERM), créé en 1971, travaille avec le Comité TC 37 de l'ISO (Organisation Internationale de Normalisation) et en étroite collaboration avec l'UNISIST à l'utilisation de l'ordinateur en terminologie et lexicographie. Pour ceux qui n'en seraient pas informés, précisons que l'UNISIST est un programme de l'UNESCO chargé spécialement d'étudier l'établissement d'un système mondial d'échanges de l'information scientifique et technique.

CLASSIFICATION DES ETRES VIVANTS

تصنيف الاحياء

Français	Anglais	1971	1967	المورد	1970	المهمل	1966	غالب	1957
règne	Kingdom				مملكة	مملكة	مملكة، عالم	مملكة، عالم، دؤجة	مملكة، عالم، دؤجة
embranchement	division				قسم	قسم	قسم	قسم	قسم
sous-embranchement	subdivision				قسم	قسم	قسم	قسم	قسم
classe	class				طائفة	طائفة	طائفة، صنف	طائفة، صنف	طائفة، صنف
sous-classe	sub-class				شئب	شئب	طائفة، صنف	طائفة، صنف	طائفة، صنف
super-ordre	super-order								
ordre	order				رتبة	رتبة	رتبة	رتبة	رتبة
sous-ordre	sub-order				رتبة	رتبة، فصيلة	رتبة، رتبة	رتبة، رتبة	رتبة، رتبة
groupe	group				مجموعة	مجموعة	مجموعة	مجموعة	مجموعة
sous-groupe	sub-group				مجموعة	مجموعة	مجموعة	مجموعة	مجموعة
super-famille	super-family				فصيلة	فصيلة	فصيلة	فصيلة	فصيلة
famille	family				فصيلة	فصيلة	فصيلة	فصيلة	فصيلة
sous-famille	sub-family				فصيلة	فصيلة	فصيلة	فصيلة	فصيلة
(= tribu)	(= tribe)				فصيلة	فصيلة	فصيلة	فصيلة	فصيلة
genre	genus				عصاة	عصاة	عصاة	عصاة	عصاة
sous-genre	sub-genus				جنس	جنس	جنس	جنس	جنس
espèce	species				نوع	نوع	نوع	نوع	نوع
sous-espèce	sub-species				نوع	نوع	نوع	نوع	نوع
variété	variety				نوع	نوع	نوع	نوع	نوع

- 1957 Dictionnaire des Termes Agricoles, Moustapha Chéhab, Le Caire (français-arabe)
1966 Dictionnaire des Sciences de la Nature, Edouard Ghalib, Beyrouth (multilingue)
1967 Dictionnaire AL MAWRID, Munir Ba'albaki, Beyrouth (anglais-arabe)
1970 Dictionnaire AL MANHAL, Jabbour Abdel Nour, Souheil Idriss, Beyrouth (français-arabe)
1971 Lexique de Zoologie, Bureau de Coordination de l'Arabisation, Rabat (anglais-français-arabe)

DICTIONNAIRES CITES

Le Petit Robert, Paul Robert, Société du Nouveau Littre, Paris, 1972.

Le Petit Larousse, Librairie Larousse, Paris, édition de 1959.

Grand Larousse Encyclopédique. Librairie Larousse, Paris, 1960-1964, suppl. 1968.

Harrap's New Standard French and English Dictionary, J.E. Mansion, revised and edited by R.P.L. Ledésert and M. Ledésert, Harrap, London, 1972.

Dictionnaire des termes agricoles, Mustapha Chéhab, Le Caire, 1957.

Dictionnaire des termes forestiers, Mustapha Chéhab, Damas, 1962.

Lisân-1-3arab, Ibn Mandhûr, Beyrouth, réédition sans date (13ème siècle).

Tâj-1-3arûs, Murtadâ Az-Zabîdî, édition de Boulaq, 1890 (17ème siècle).

Matn-1-Lugha, Ahmed Ridâ, Beyrouth, 1958.

Al Wasît, Académie du Caire, Le Caire, 1960.

Al Munjid, Père Louis Ma3lûf, Beyrouth, 1908, et rééditions.

OUVRAGES CITES

Bibliography of French/English/French and polyglot dictionaries, Vojto Margaret, Translation Center, Carnegie Mellon, USA 1975.

Méthodologie Générale de l'Arabisation de Niveau, Ahmed Lakhdar-Ghazal, Rabat, 1976, 2ème éd.

ORGANISATIONS CITES

ALECSO (Arab League Educational, Scientific and Cultural Organization) Organisation de la Ligue Arabe pour l'Education, la Science et la Culture, Le Caire, Egypte.

Bureau de Coordination de l'Arabisation, Rabat, Maroc.

CASTARAB Conférence des Ministres des Etats Arabes, chargés de l'application de la Technologie au Développement, Rabat, août 1976.

ESRIN European Space Research Institute (Institut Européen de Recherche Spatiale) dépendant de l'Agence Spatiale Européenne, Frascati, Rome, Italie.

INFOTERM Centre International d'Information sur la Terminologie, Vienne, Autriche.

ISO International Standardization Organization (Organisation Internationale de la normalisation), Genève, Suisse.

PNUD (UNDP) Programme des Nations Unies pour le Développement, New York, USA.

UNESCO, Paris, France.

UNISIST Programme de l'UNESCO pour l'établissement d'un système mondial d'échange de l'information scientifique et technique.

LEXAR

IDENTIFICATION (BUNGLE) 4.01 78

L 1.2.3

	DATE	SEC
CONTROLE		
BASE		
VERIFICATION		

9 3

5

7

13

15

18

00

00

no session

out

E

I

C

E

I

C

20	23	25	30	35	40	45	50	55	60	64
0001										
0101										
0102										
0103										
0201										
0202										
0203										
0301										
0302										
0303										
0304										
0305										
0306										
0307										
0308										
0309										
0310										
1001										
1101										
1102										
1103										
1201										
1202										
1203										
1301										
1302										
1303										
1304										
1305										
1306										
1307										
1308										
1309										
1310										