

***A propos d'un Dictionnaire Electronique
Arabe/Anglais /Français
dans le Domaine des Hydrocarbures***

**BOUHADIBA Farouk
ROUABHI Miloud
(Université d'Oran)**

Résumé :

Cet article présente une idée globale sur les dictionnaires électroniques. Celle-ci est suivie d'une description succincte de ce type de dictionnaires et de leur utilisation. Cette description nous permettra de présenter un petit échantillonnage d'un dictionnaire électronique dans le domaine des hydrocarbures dans un premier temps, et d'un futur dictionnaire électronique dans le domaine du tourisme dans le cadre d'un autre article.

Notre souci majeur pour l'élaboration d'un dictionnaire électronique arabe-anglais-français dans le domaine des hydrocarbures est dicté principalement par une demande de plus en plus croissante que nous avons constatée auprès des utilisateurs dans ce domaine. Ces derniers (Cadres, Ingénieurs, Techniciens, etc.) reçoivent des documents rédigés en Anglais concernant soit des caractéristiques techniques d'une pièce de rechange (Mud pump, gas turbine, crown block, drill pipes, etc.), soit des prospectus et autres documents dans le domaine du startup (procédures de démarrage, ...), de

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

l'exploration (analyses de données géologiques, sismiques, cartes, graphes, etc.), de la production et du développement (forage, estimation des réserves en pétrole ou en gaz), ou bien dans le domaine de la production, du « process », des études quantitatives et qualitatives (reserves estimation, uncertainty, water saturation, decline curve analysis, etc.). Ces utilisateurs maîtrisent dans une certaine mesure les rudiments de la langue anglaise mais ils se trouvent confrontés à deux problèmes majeurs :

- a) Soit tenter de traduire et de comprendre le document à l'aide d'un dictionnaire papier.
- b) Soit faire appel à un traducteur humain qui, parfois, ne maîtrise pas le domaine des hydrocarbures.

Dans les deux cas de figure, il en résulte une perte de temps considérable, surtout lorsqu'il s'agit de répondre à un document (fax, e-mail, ou autre) dans des délais très réduits, une traduction parfois erronée qui est beaucoup plus due au foisonnement de termes techniques que le document comporte qu'aux structures syntaxiques et autres de la langue dans laquelle le document est rédigé (Anglais Technique des Hydrocarbures (ATH)).

Ainsi, l'idée a germé dans notre Laboratoire de Recherche (LDLD, Université d'Oran, Algérie) de s'intéresser au Traitement Automatique des Langues Naturelles (TALN) en général et à celui de la Langue Arabe en particulier. Cette dernière se faisant présente à petits pas dans les autoroutes actuelles de l'information.

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

La conception et la réalisation du dictionnaire en question est certes à ses débuts. Elle devrait prendre en considération les propriétés lexicales, morphologiques, syntaxiques et sémantiques de chaque item stocké dans la base de données afin d'optimiser la traduction des termes dans le domaine des hydrocarbures.

1. Introduction

Le monde de l'informatique est sûrement l'un des milieux qui évoluent de plus en plus vite. En effet, plusieurs concepts voient le jour. Le plus récurrent étant celui de la veille technologique. Celle-ci représente un concept assez récent qui désigne les activités de surveillance de l'environnement technologique afin de capter toutes les informations stratégiques. La réalisation d'un système d'aide à la veille est basée sur l'analyse sémantique du contenu et les documents acquis contiennent des informations contrôlées sur les acteurs de l'innovation et sur l'avancée des recherches pour tous les secteurs.

Les documents les plus riches en information technologique sont les brevets. Ces derniers contiennent en effet des informations contrôlées sur les acteurs de l'innovation et sur l'avancée des recherches pour tous les secteurs (la plus part du temps, les documents sont récupérés au format électronique et peuvent ainsi être appliqués directement par des logiciels). Pour le cas des dictionnaires, le spécialiste doit analyser plusieurs références du dictionnaire.

- Dictionnaire et hypertexte

Comme point de départ, partons du terme « livre électronique ». Ce terme nous semble restrictif dans le sens où le livre désigne un support particulier de l'écrit. De plus, il nous paraît assez restrictif de parler de livre là où tous les supports de l'écrit, du son et de l'image sont convoqués. Enfin, le terme nous semble inopportun car la juxtaposition des deux mots « livre » et « électronique » semble, à première vue, comme antithétique: le livre désigne d'abord le support physique de l'écrit. Livre dérive du Latin « liber » qui désigne en fait l'aubier, l'arbre dont on tire les feuilles sur lesquelles on écrit. Or le mot « électronique » superpose à l'objet matériel initial, le livre de papier que tout le monde connaît, un nouvel objet immatériel défini par un ensemble de procédures d'accès et une structuration logique. A cette signification restrictive et antithétique du terme « livre électronique » s'ajoute un statut hypothétique: le livre électronique n'est pas encore réalisé et quand bien même il le serait, son statut demeurerait encore largement indéterminé. Certains croient l'avoir déjà rencontré dans les supermarchés sous la forme de cd rom, d'hyper livres ou sur internet. Néanmoins, en regardant de près, il semblerait que le support numérisé des livres classiques serait confondu avec le livre électronique pour deux raisons :

- D'un côté il y aurait un mode de présentation particulier qui équivaldrait fonctionnellement à l'impression classique ou plus exactement à une accumulation de pages imprimées.

- D'un autre côté, les procédures traditionnelles d'accès à l'écrit ainsi qu'aux différentes modalités de

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

connaissance (images fixes, animées, sons) seraient susceptibles d'être modifiées, au point de poser des problèmes d'ordre cognitif inconnu jusqu'à présent. Dans ce cadre, nous voyons apparaître de nouveaux objets encore mal définis, caractérisés de façon imprécise comme étant à la fois des supports physiques d'informations et des opérations de manipulation. S'agit-il déjà de livre électronique? Il convient de savoir sous quelle(s) forme(s) le livre se présentera, si le livre électronique sera le complément ou le substitut du livre de papier et dans l'une ou l'autre éventualité, quel serait son rôle pour l'utilisateur. Une question à laquelle nous n'avons pas de réponse pour le moment.

2. Une brève définition d'un dictionnaire électronique

Commençons d'abord par la définition classique d'un dictionnaire. Celui-ci réfère généralement à un recueil par ordre alphabétique ou autre des mots d'une langue, avec leur explication (lexique, vocabulaire, glossaire), etc. Par contre, le dictionnaire électronique est un dictionnaire de travail sur support informatique (disque dur ou CD ROM). Les fonctionnalités varient selon les éditeurs des dictionnaires associés à des logiciels TAO (Traduction Assistée par Ordinateur). Ils permettent de reconnaître des termes source enregistrés dans le dictionnaire et d'en proposer automatiquement une traduction grâce à une reconnaissance automatique. Cette fonction permet d'assurer la consistance terminologique tout au long du texte.

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

2.1 Caractéristiques d'un dictionnaire électronique

En termes généraux, un dictionnaire électronique est caractérisé par :

- Un accès thématique, une multiplication d'indexes, des recherches en « texte intégral », c.-à-d. de toutes les occurrences d'un mot dans l'ensemble d'une page, d'un document ou d'un ouvrage.
- Une navigation de type hypertexte y est proposée: certains mots rencontrés dans une définition renvoient à d'autres rubriques (notion de liens), à un approfondissement ou à une illustration (cartes et citations)

Plusieurs dictionnaires électroniques dans différents domaines existent de nos jours. Citons, quelques exemples:

2.2 Quelques exemples de dictionnaires multilingues

- Le Franglophile (Scott Carpenter et Philippe Caudron) : Ce dictionnaire bilingue français-anglais informatisé compte plus de 100,000 termes ainsi que des dizaines de milliers d'entrées secondaires. Il est distribué sous forme de shareware.
- Freetranslation.com : Service de TA de textes (Anglais/Français, Français/Anglais (A/F, F/A)).
- Systran : Le moteur de recherche Alta Vista qui donne accès à un service de traduction automatique de textes (A/F, F/A).

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

- TransSearch (Université de Montréal, Québec)
TransSearch permet de traduire et de mettre en contexte des termes de l'anglais au français et vice versa.

Il serait peut être utile de mentionner d'autres types de dictionnaires électroniques plus récents et mieux adaptés à notre recherche. Nous citerons par exemple:

- <http://raaube.cegepbceapp.qc.ca/Internet.htm>: Ce dernier constitue un ensemble de dictionnaires spécialisés (français / anglais) élaborés par l'Université de Sherbrooke. Nous y trouvons également des encyclopédies spécialisées dans les domaines, entre autres, de la géologie et de la géographie qui nous intéressent particulièrement.
- <http://www.dicorama.com>. représente un ensemble de dictionnaires avec un moteur de recherche. Cet outil de recherche permet d'effectuer des recherches dans plus de 670 dictionnaires. Il propose des traducteurs en ligne (environ 250) permettant de traduire gratuitement dans plus de 40 langues. Ceci dit, les traductions qu'il propose ne sont pas tout à fait fiables. Le domaine de la technologie nous paraît bien élaboré.
- <http://www.granddictionnaire.com> est spécialement conçu comme étant le Grand Dictionnaire Terminologique qui représente à notre sens un site unique dans son genre puisqu'il comprend un fonds terminologique d'environ trois millions de termes techniques en français et en anglais dans un nombre très varié de domaines de spécialité.
- <http://tis.consilium.eu.int> est un site du Conseil de l'Union Européenne (Council of the European Union)

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

basé sur le système TIS (Terminological Information System). Sa banque de données terminologiques comprend les abréviations et la terminologie utilisées au niveau du Conseil de l'Union Européenne.

- <http://www.logos.it/query/query.html> nous renvoie vers le Logos Group qui propose une base de données terminologiques multilingue. L'intérêt de ce groupe est, nous semble-t-il, les équivalences qu'il propose pour une terminologie dans plusieurs langues d'Europe y compris en chinois, en japonais et même en Latin.
- <http://www.admin.ch/ch/f/bk/sp/dicos/biling.html> nous semble intéressant dans le sens où en plus des dictionnaires bilingues et multilingues qu'il propose, il offre également des dictionnaires spécialisés dans les domaines de l'industrie, de la chimie et de la physique qui peuvent nous servir pour notre propre conception d'un dictionnaire électronique arabe anglais français dans le domaine des hydrocarbures où ces domaines (technique, industrie, chimie, physique, géologie) sont très récurrents.
- <http://www.softissimo.com/products/infotrad.htm> est un site très connu pour les offres en ligne qu'il propose. Celles-ci incluent des dictionnaires électroniques tels que le Robert & Collins bilingual French/English-English/French disponible également sur CD-ROM ou bien le dictionnaire EuroDico qui est un dictionnaire électronique dans plusieurs langues européennes (06) telles que l'anglais, le français, l'espagnol, etc.

Nous ajouterons à cette liste des dictionnaires de traduction développés dans des centres de recherches qui s'intéressent de plus en plus au Traitement Automatique

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures
des langues dites orientales et de langue arabe. Citons
comme exemple :

- Le Dictionnaire Multicodict (Français, Arabe, Anglais). Ce dictionnaire de traduction des termes informatiques et autres a été réalisé au Centre de recherche Lucien Tesnières à Besançon (France).
- Le Traducteur Technique : Dictionnaire technique gratuit en ligne pour l'aide à la traduction d'expressions techniques en Français, Anglais, Allemand et en Espagnol.
- Le Programme Technolange qui est un programme français dans le domaine des technologies de la langue et qui a été lancé en 2002 pour la création de ressources linguistiques et logicielles, pour la normalisation et la création de portails Web pour diffuser des informations dans le domaine des technologies de la langue et où la langue arabe y est analysée.¹

Il n'en demeure pas moins que ces sites, centres de recherche, et autres ne proposent pas ou peu de ressources dictionnairiques pour la langue arabe. Ceci est dû au fait que cette langue, en dépit de son entrée à l'ONU comme langue officielle depuis 1974, elle demeure peu exploitée malgré ses systèmes morphosyntaxiques, lexicaux et autres qui la disposent au Traitement Automatique. Cet état de fait est dû beaucoup

¹. Voir à ce sujet S. CHAUDIRON, Université de Lille 3, « *La place de l'arabe dans le programme Technolange* », RML 5, pp. 15-40, Dar El Gharb Ed. 2007.

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

plus à ses caractéristiques propres de langue non Indo-européenne avec des procédés de non enchaînement et son système de vocalisation qui est très différent des langues généralement analysée et mises en ligne.

A ce titre, A. Ghénimi synthétise ces difficultés d'analyse et d'informatisation de cette langue en ces termes :

« La langue arabe, qui constituait le centre d'intérêt de la traduction et des traducteurs pendant des siècles dans le passé, n'a pas eu cette chance avec les nouvelles technologies. Ses spécificités morphologiques et syntaxiques constituent un obstacle, jusqu'à présent, insurmontable. Très peu de travaux ont été réalisés dans la traduction automatique du ou vers l'arabe, et encore moins entre l'arabe et le français. »²

Cette vue d'ensemble des fonctionnements et utilisations de ces ressources dictionnaires et moteurs de recherche nous permettrons d'aborder une architecture du dictionnaire électronique dans le domaine des hydrocarbures sur lequel nous travaillons actuellement.

Il serait intéressant de mentionner que des recherches se font de plus en plus présentes dans le cadre de la modélisation de la langue arabe et ce à base de modèles proposés pour d'autres langues mais adaptés à cette langue en question. Nous citerons, à titre d'exemple et

². GHENIMI Abdelouafi, Conception d'un modèle de traduction arabe-français appliqué au domaine des mathématiques. Thèse de Doctorat, Centre lucien Tesnière, Besançon 2007, p. 8.

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

sans être exhaustifs, le cas du modèle le plus connu qu'est celui du système Sakhr (**Ajeeb**) que l'on peut consulter sur <http://lexicons.ajeeb.com/Results.asp>, **Kalimet** sur <http://www.kl28.com/lesanalarab.php>, ou bien **LisanElArab/Al-Qamous-Al-Mouhit** que l'on retrouve en ligne sur <http://www.content.com.sa/Languages/>. Les systèmes adoptés sont principalement TEI (Text Encoding Initiative)³ et LMF (Lexical Markup Framework)⁴, le système d'interrogation ADIQTQ (Arabic Dictionary Query TOols)⁵, ou bien le système MORPHE de John, R.R. Leavitt⁶

3. Les phases de construction du dictionnaire

Tout d'abord il faut préparer le lexique concerné et classer les items par catégorie selon les domaines (pétrole et hydrocarbures) en trois langues. Cette étape a déjà été entamée par la création d'une base de données lexicale partant de l'anglais, vers le français puis vers l'arabe. Cette chronologie ayant été dictée par le fait que la terminologie existante dans ce domaine particulier se trouve principalement en Anglais et dans une moindre mesure en français. Quant à la langue arabe, nous avons

³ . Sperberg-Mcqueen & Burnard, 2005.

⁴ . Francopoulo & George, 2008

⁵ . cf. F. BACCAR, A. KHEMAKHEM, B. GARGOURI, K. HADDAR, et A.BEN HAMADOU : « *Modélisation normalisée LMF des dictionnaires électroniques éditoriaux de l'arabe* »

⁶ . John, R.R. Leavitt, MORPHE: A Practical Compiler for Reversible Morphology Rules. Center for Machine Translation, Carnegie Mellon University, Pittsburg, USA).

été souvent obligés de faire dans la paraphrase du fait que les dictionnaires dans ce domaine sont pratiquement inexistantes et du fait que les dictionnaires classiques de la langue arabe se présentent soit par schème, par racine, ou parfois même par dérivation.⁷

La deuxième étape de ce dictionnaire a été de remplir la base de données et de préparer les règles pour chaque mot du dictionnaire pour obtenir une base de connaissances puis construire autour de ce module principal les modules d'entrée et de sortie (les interfaces). Nous avons pensé à un modèle de données et une interface du lexique structuré en système hypertexte. Cette structure doit être homogène sur le plan lexicosyntaxique et sémantique pour permettre une meilleure recherche, d'où la nécessité d'avoir une bonne représentation des connaissances avec des méthodes d'accès et des dispositifs de stockage pour un éventuel enrichissement de la base par une terminologie spécialisée dans le domaine des hydrocarbures. Ce système doit avoir également une structure flexible pour procéder à son extension.

⁷. Cf. l'article de F. BACCAR, A. KHEMAKHEM, B. GARGOURI, K. HADDAR, et A.BEN HAMADOU : « *Modélisation normalisée LMF des dictionnaires électroniques éditoriaux de l'arabe* » dans lequel ces auteurs relèvent des problèmes d'ordre structurel des dictionnaires de la langue arabe tels que Al-Ain 'العين', Tahdhib Al-Loghah 'تهذيب اللغة' (classification à base phonétique), Lisan-El-Arab 'لسان العرب', Qamous-Al-Mouhît 'قاموس المحيط' (classification alphabétique), soit par racine comme dans Al-Wassit 'الوسيط' ou bien Al-Mosbah Al-Mounir 'المصباح المنير'.

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

4. La phase de conception du dictionnaire électronique arabe/ anglais/ français dans le domaine des hydrocarbures

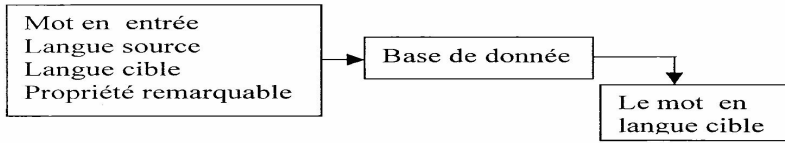
Cette phase a pour but de détailler l'ensemble des démarches suivies et les choix effectués tout au long de la conception du dictionnaire. Nous utilisons une méthode descendante. Nous présenterons l'architecture globale du système, ainsi que les différentes étapes suivies pour la conception de chaque module de ce dernier. Au niveau du **découpage** nous segmentons la phrase en unités lexicales qui seront utilisées par la suite. Au niveau **lexical** nous présentons l'ensemble des modules de l'analyse lexicale, la base de connaissance, le gestionnaire de l'analyse lexicale, les opérations de base du dictionnaire (suppression, mise à jour, modification etc.). Au niveau **syntaxique**, la grammaire adoptée est la table **LR**. Au niveau sémantique nous exposons les propriétés remarquables. Cette phrase se termine par la description des erreurs qui peuvent se présenter pendant cette analyse.

5. Le fonctionnement du système

5.1. L'entrée : Mot

Pour récupérer la traduction d'un mot simple en anglais, en français ou bien en arabe, il suffit de lancer la recherche dans la table de données sans prendre en considération l'information lexicale, syntaxique de ce mot.

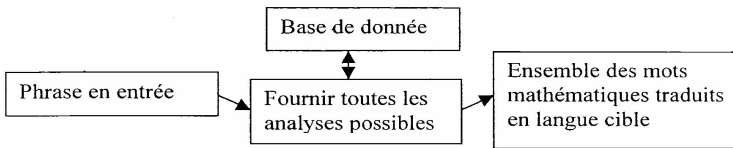
A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures



Entrée Mot

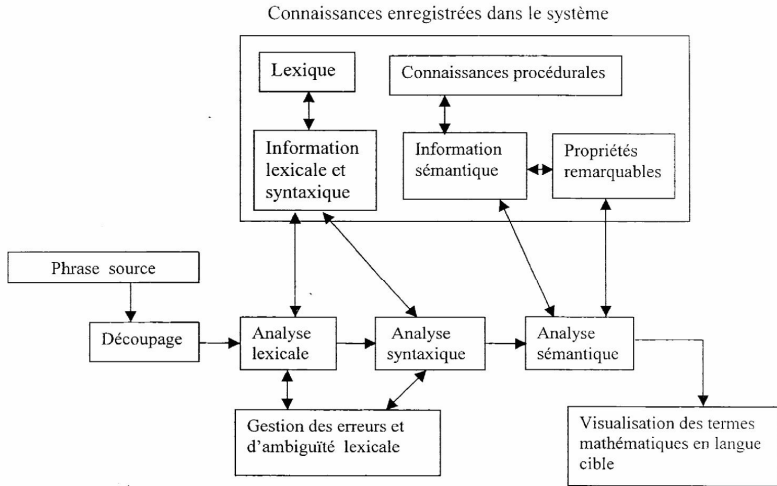
5.2. L'entrée: Phrase

Pour avoir la traduction d'une phrase, il faudrait analyser cette dernière en entrée, détecter les erreurs lexicales et retourner certaines informations de chaque mot en Anglais (langue pivot), ainsi que les informations lexicales et syntaxiques.



Entrée Phrase

5.3 L'architecture fonctionnelle



L'architecture fonctionnelle du dictionnaire

6. Conceptions des différents modules

6.1 Découpage

Le découpage est une phase qui prépare l'analyse lexicale en découpant la phrase en entrée séparée par des blancs ou par des signes de ponctuation (point, apostrophe) en mots. Chaque mot de la phrase, ainsi que les informations relatives à ce dernier doivent figurer dans la base de connaissances pour pouvoir comparer les mots de la phrase avec ceux existants.

Pour avoir un meilleur résultat du découpage, nous avons utilisé un vecteur pour regrouper tous les mots

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures
constituants de la phrase en entrée. Toute Chaîne P de
mots M est une suite finie ordonnée.

$$\mathbf{P} = M_1 M_2 M_3 \dots M_n$$

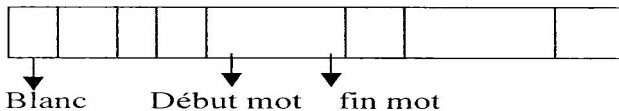
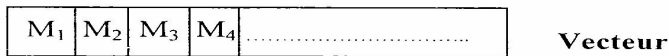
Avec : M_i en tant que mot de rang i dans $\mathbf{P}$

Ensuite nous traitons les étapes parcourues dans ce module selon les étapes ci-dessous :

Parcourir la phrase en entrée.

1. Si le caractère de position est (i) (blanc ou apostrophe), nous passons au caractère de position (i+1). Si ce dernier est un blanc ou une apostrophe nous passons au suivant et ainsi de suite jusqu'à ce que le caractère soit différent du blanc ou de l'apostrophe. Dans ce cas, le début mot reçoit ce caractère. Ensuite, le contraire est fait pour déterminer la fin mot et nous plaçons le mot dans un vecteur. La même procédure se répète pour tous les mots de la phrase.

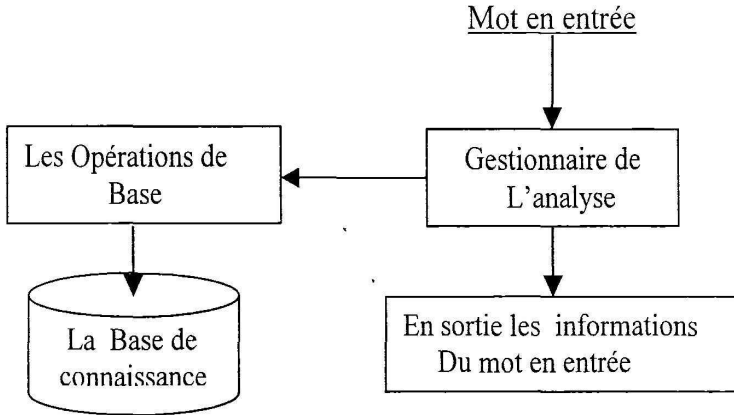
2. Finalement l'on obtient le vecteur qui contient tous les mots de la phrase.



Exemple de découpage

6.2 Le module lexical

Le module lexical a pour but de tirer toutes les informations grammaticales concernant un mot. Ces informations peuvent être nécessaires par la suite au niveau de l'analyse syntaxique.



L'analyseur lexical

L'analyse lexicale comme le montre la figure ci-dessus est constituée de trois modules :

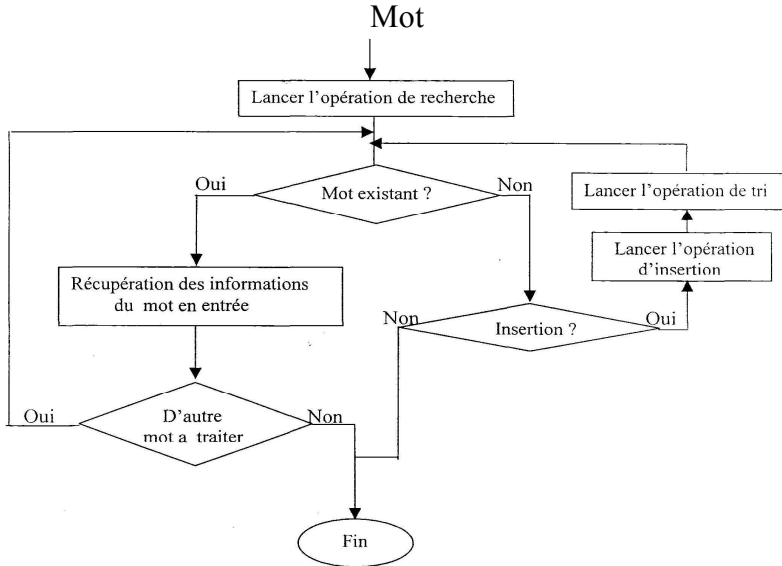
- Un Gestionnaire de l'analyse.
- Une Base de connaissance.
- Opération de base de dictionnaire.

Le langage naturel est très riche et ne cesse d'évoluer. Pour s'adapter au maximum à ces contraintes, nous avons

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures
conçu trois modules qui rendent l'analyse lexicale plus
souple.

6.2.1 Le gestionnaire de l'analyse lexicale

Le gestionnaire de l'analyse est un programme défini par le concepteur pour ordonner les différentes opérations principales de l'analyse lexicale (c'est-à-dire l'accès au dictionnaire en lançant une opération comme la recherche, la récupération des informations, etc.). La figure présente le diagramme du gestionnaire



Le gestionnaire de l'analyse lexicale

6.2.2 Les opérations de base

- **Recherche** : La recherche d'un mot dans le dictionnaire se fait selon le diagramme suivant.

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

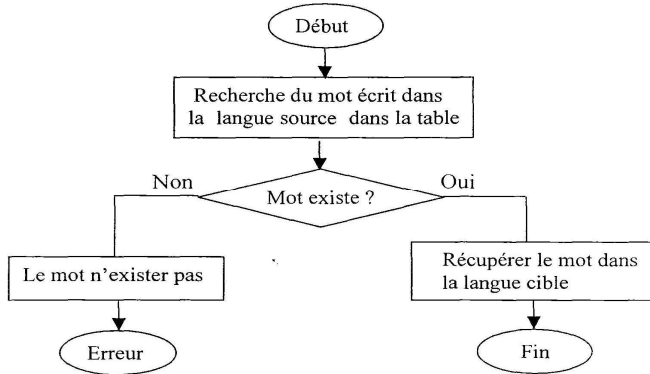


Schéma représentant la recherche d'un mot dans le dictionnaire

- **L'insertion** : Cette opération est lancée pour insérer un nouveau mot dans la base de connaissances avec ses informations (la traduction du mot en trois langues « anglais français, arabe », la catégorie grammaticale (sujet, nom, adj. etc.), le genre, le nombre et les propriétés remarquables.

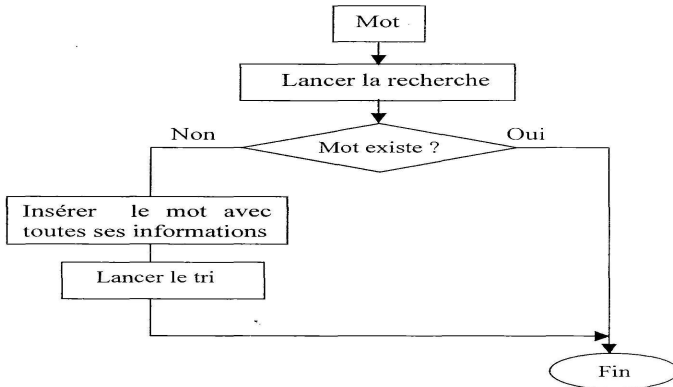


Schéma représentant l'insertion d'un mot dans le dictionnaire

- **La suppression** : Cette opération de supprimer un mot de la table de données avec toutes ses informations enregistrées.

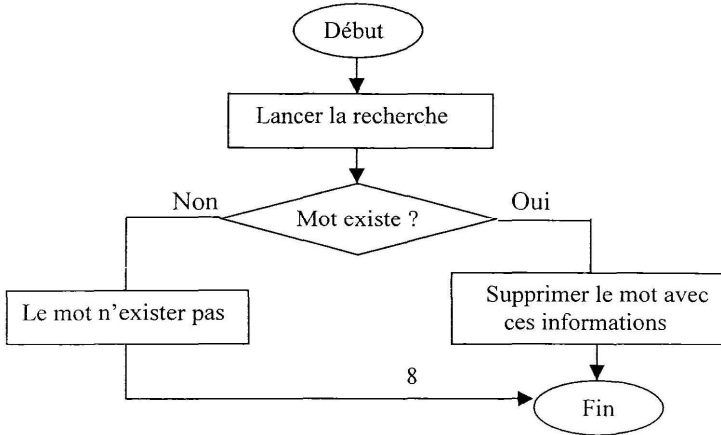


Schéma représentant la suppression d'un mot dans le dictionnaire

- **Le tri** : le tri du dictionnaire est très important. Il facilite la recherche d'un mot dans le dictionnaire.

- **La modification** : cette opération consiste à mettre à jour les informations du dictionnaire.

6.2.3 La Base de connaissance

La base de connaissances constitue un outil indispensable pour le système. Elle contient des informations morphosyntaxiques et sémantiques relatives aux mots.

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

6.2.3.1 Les connaissances utilisées dans l'analyse lexicale et syntaxique

- Informations lexicales et syntaxiques

Les connaissances lexicales et syntaxiques utilisées sont: les catégories grammaticales du mot (nom, verbe, adj., sub. etc.), le nombre (singulier ou pluriel) et le genre (féminin ou masculin) des noms.

6.2.3.2 La structure de la base de données

La base de donnée du dictionnaire est constituée de plusieurs tables Une table des mots (hydrocarbures), une autre pour les propriétés remarquables etc. On réalise ainsi les liens entre ces tables par des relations. La récupération des données se fait par l'utilisation des requêtes SQL (Select, Jointure, etc.).

- La structure de la table des mots

Num	Français	Genre	Nombre	Catégorie

Mot:

On utilise des différents mots de plusieurs domaines (Forage, analyses sismiques, process, production, etc.)

Genre : Féminin, Masculin

Nombre: Singulier Pluriel

Catégorie: Nom, verbe, déterminant, adjectif, article, substantif

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

Num	Français	Genre	Nombre	Catégorie
01	la	féminin	singulier	article
02	fonction	féminin	singulier	nom
03	convergente	féminin	singulier	adjectif

La fonction convergente

6.3 Module syntaxique

Le but de l'analyse syntaxique est de trouver d'abord si la phrase en entrée est correctement formée ; c'est-à-dire, vérifier l'agencement des mots dans la phrase suivant une grammaire bien définie, il s'agit d'associer à chaque mot de la phrase sa catégorie grammaticale parmi celles qui lui ont été assignées par l'analyseur lexical.

Exemple : « *La probabilité est une fonction croissante* »

La probabilité = GN

{La =— article, Probabilité= substantif}

est une fonction croissante = **GV**

{Est = verbconj}

Une fonction croissante = **COD**

{Une = déterminant, Fonction= substantif, Croissante = adj}

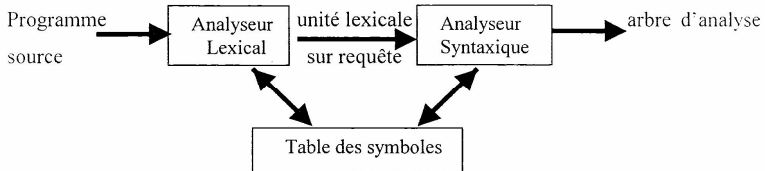
6.3.1 Le rôle de l'analyse syntaxique

Le rôle général de la partie frontale est

- Activer l'analyseur lexical.
- Vérifier la conformité syntaxique.
- Construire l'arbre d'analyse.

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

- Préparer ou anticipe la traduction.
- Gérer les erreurs communes de syntaxe.



Le rôle de l'analyseur syntaxique

6.3.2 La grammaire

Pour analyser une phrase du point de vue syntaxique, il est nécessaire d'utiliser une grammaire qui décrit la structure des suites des mots dans un langage particulier. En connaissant une grammaire, l'analyseur peut affecter un arbre syntaxique à la phrase. Le but essentiel de notre grammaire est de générer une phrase.

Comme nous le savons, la langue française admet plusieurs types de phrase comme toutes les autres langues (déclaratives, interrogatives, exclamatives, etc.). Nous considérons seulement les phrases déclaratives simples du type **SVO**.

Suite à une étude détaillée avec des linguistes, nous sommes arrivés à concevoir la grammaire formelle suivante:

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

N : Les entités en majuscules.

Σ : Les entités en minuscule.

S : est l'axiome

S	<u>1</u> →	GN GV
GN	<u>2 3 4</u> →	SUJET COMP / pronper / inf
SUJET	<u>5</u> →	DET NOM ADJ
DET	<u>6 7 8 9</u> →	article / pronpos / pronodem / vide
NOM	<u>10</u> →	subs
ADJ	<u>11 12</u> →	adj / vide
COMP	<u>13 14</u> →	LIEN SUJET / vide
LIEN	<u>15 16 17</u> →	prep / conj / vide
GV	<u>18</u> →	verbconj PART COD
PART	<u>19 20</u> →	part / vide
COD	<u>21 22</u> →	SUJET COMP / vide

Signification des terminaux de la grammaire

Terminal	signification
Pronper	pronom personnel
Inf	infinitif
Art	article
Pronpos	pronom possessif
Prondem	pronom démonstratif
Vide	aucune action

Adj	adjectif
Subs	substantif
Prep	préposition
Conj	conjonction
Verbconj	verbe conjugué
Part	participe

Significations des non terminaux de la grammaire

Non terminal	Signification
SUJET	Sujet de la phrase
COMP	Complément du sujet
GN	Groupe nominal
COD	Complément d'Objet Direct
GV	Groupe verbal
PART	Participe
NOM	Nom
ADJ	Adjectif
DET	Déterminant
LIEN	Pour donner une préposition ou une conjonction.

6.3.3 La Méthode LR

L: Signifier parcours de la chaîne d'entrée de grammaire vers la droite «left to right scanning of input»

R : dérivation droite inverse en parcourant la chaîne

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures
d'entrée «constructing a right most derivation inverse»

67.3.3.1 Définition

Soit $G = (N, \Sigma, P, S)$ on dira $[A \rightarrow B_1 B_2, u]$ est un article (item) LR(K) ; si $A, B_1 B_2 \in P$ et $u \in \Sigma^{*k}$, $|u| \leq \Sigma$, u est appelé look head, $B_1 B_2 \in (N \cup \Sigma)$, $A \in N$.

Closure

Procédure closure (I)

Début :

Répéter pour chaque article $[A \rightarrow \alpha . B \delta, u] \in I$ et

chaque production $B \rightarrow \gamma \in P$, / $[B \rightarrow \gamma, \text{First}_k(\delta u)] \in I$

Faire

Ajouter $[B \rightarrow \gamma, \text{First}_k(\delta u)]$ jusqu'à ce que l'on ne puisse rien ajouter à I

Fin

$A, B \in N_i \quad \alpha, \gamma, \delta \in (N \cup \Sigma)^*, u \in \Sigma^{*k} \cup \{\}$

6.3.3.2 La table LR et la pile LR

Les tables LR servent à appliquer les règles de la grammaire par une stratégie ascendante. Les entrées de la table sont une valeur grammaticale et un état qui est toujours le sommet de la pile LR.

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

Valeurs grammaticales possibles

.....	substantif	inf	adjectif		S	A	
0							
1							
2							
.							
.							

Etats
goto

Partie action

Partie

Schéma représentant les entrées de la table LR

Pourquoi ce choix?

- Le choix de la méthode d'analyse a été la méthode ascendante LR parce qu'elle est plus déterministe pour la gestion des erreurs que la méthode descendante.

- Simple à implémenter

6.3.3.3. Construction de la table d'analyse LR

Soit $c = \{I_1, I_2, \dots, I_n\}$ une collection d'ensemble d'articles de chaque I_j représente l'état J.

a) construction de la partie **Action**

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

Si [$A \rightarrow \alpha .aB, u$] $\in I_j$] et goto (I_j, a) = I_k alors **action** (j, a) = d_k

Si [$A \rightarrow \alpha , u$] $\in I_j$] alors action (j, u) = y ou y est le n° de la règle de production $A \rightarrow \alpha$

Si [$S \rightarrow S, \$$] $\in I_j$ alors **action** ($j, \$$) = accept.

b) construction de la partie **goto**

Goto (I_j, A) = I_k alors dans I table Goto (j, A) = k

On note que l'état initial de l'analyseur est l'état qui contient l'article $S' \rightarrow S, \$$ qui s'ajoute à la grammaire précédente.

7. Un Echantillon de la Base de données Lexicales

A.A.O.D.C.: (Engl) : American Association of oil well drilling contractors.

(Fr) Association américaine des entrepreneurs de forage des puits pétrolifères

(Ar) الجمعية الأمريكية للمتعهدين في حفر آبار البترول

Abate (to) (Engl) : (Fr) Diminuer : (Ar) نقص, قلل

Abnormal (Engl) : (Fr) Anormal : (Ar) خالف للمؤلف ، غير عادي

Abnormal dip (Engl) : Pendage anormal (Fr): (Ar) غير

Abnormal Stress : Contrainte anormale ضغط غير عادي

: عادي

Above ground : A la surface : على سطح

: à ciel ouvert : على سطح مكشوف

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

- : بلاطة المنجم : carreau d'une mine
- Abradant : كاشط، حكاك، سحاج : Abrasif
- Abrasion of sucker rods on tubing : Usure des tiges de
pompage dans les tubes : إستهلاك سيقان الضخ في الأنابيب
- Abreast : مجابهة، مواجهة، على الخط نفسه : de front
- :
Abreast connection : تزويج على Couplage en parallèle
التوازي :
- Babcock tube: أنبوبة إختبار : Fiole d'essai type « Babcock »
- Back : ظهر : Dos :
Back نوع " بابكوك :
: خلف : Arrière :
: شاشة : Ecran :
: في الخلف : en arrière :
: في العودة : en retour :
Back axis : جسر خلفي : Pont arrière
- Back balance : Soutènement provisoire du puits :
مؤقت للبر
- Back crank pumping : ضخ (Prod.) : Pompage combiné (Prod.) :
ممزوج
- Back deep : حفرة خلفية : Arrière-fosse
- Back digger : Retrocaveuse : (Ar. ???)
- Back draft, back draught : Refoulement : (Ar. ???)
- Back entry : ممر : Passage d'aérage (Mine) :
- التهووية (منجم)
- Back electromotive force : Force contre-électromotrice :
قوة ضد قياس القوة الكهربائية
- Cam shaft bearing : ضام : palier de l'arbre à cames :
جذع الكامات
- Car bearing : ضام : palier de wagon :
- Collar bearing : ضام الضيق : palier à collet :
- Connecting rod bearing : وسادة الساعد : coussinet de bielle :

A propos d'un Dictionnaire Electronique Arabe/Anglais /Français
dans le Domaine des Hydrocarbures

Crank-pin bearing : manetton (Ar. ???)

Double row ball bearing : roulement à double rangée de billes : مدرجة ذات صفين الكريات

Finite bearing : palier de largeur finie : طمام بعرض منتهي

Floating ring bearing : palier à coussinet flottant : طمام بوسادة طافية

Fluid bearing : palier fluide : ضمّام مائع

Guide bearing : palier guide : ضمّام موجّه

Lower guide bearing : palier guide inférieur : طمام موجّه سفلي

Upper guide bearing : palier guide supérieur : طمام موجّه علوي

8. Conclusion

Cette étude a été faite pour avoir une vision globale sur les dictionnaires existants et nous permettre de réfléchir et de proposer une étude détaillée d'un type de dictionnaire en l'occurrence un dictionnaire multilingue des hydrocarbures avec son architecture et son fonctionnement. Parmi les problèmes rencontrés, nous signalons la consultation d'un vaste ensemble de documents et l'identification des informations technologiques. Généralement, ces tâches sont assurées par un serveur d'investigation spécialisé dans la veille technologique.

Références

- GOIJON, B. : « Réalisation du système informatique VIGITEXT », Thèse de doctorat, Université de Paris 4.
- CHAUDIRON, S. « La place de l'arabe dans le programme Technolanguage », RML 5, pp. 15-40, Dar El Gharb Ed. 2007.
- GHENIMI A., « Conception d'un modèle de traduction arabe-français appliqué au domaine des mathématiques ». Thèse de Doctorat, Centre Lucien Tesnière, Besançon 2007.
- McQUEEN, S. & BURNARD, C.M.. Guidelines for Electronic Text Encoding and Interchange, TEI P5 – TEI Consortium. January 2005.
- FRANCOPOULO, G., GEORGE , M. (2008). Language resource management – Lexical mark-up framework (LMF).ISO/TC 37/SC 4 Rev. 15.
- BACCAR, F., KHEMAKHEM,A., B., GARGOURI, HADDAR, K. & BEN HAMADOU, A. :
« Modélisation normalisée LMF des dictionnaires électroniques éditoriaux de l'arabe » TALN 2008, Avignon, 9-13 juin 2008.
- LEAVITT, John, R.R. MORPHÉ: A Practical Compiler for Reversible Morphology Rules. Center for Machine Translation, Carnegie Mellon University, Pittsburg, USA).
- **Webographie** :
- Sites Internet mentionnés dans le texte.