Zeitschrift Übersetzung und Sprachen مجلة الترجمة واللغات
ISSN : 1112-3974

Les Dictionnaires Électroniques
Electronic Dictionaries

Ghenimi Abdelouafi
Université d'Oran – Algérie
g.abdelouafi@yahoo.fr

To cite this article:
Ghenimi, A. (2002). Les Dictionnaires Électroniques. *Revue Traduction et Langues 1 (1),* 86-92.

Résumé: *Cet article examine, en s'appuyant sur des exemples concrets, la problématique de la traduction automatique et de la conception des dictionnaires électroniques. L'étude a montré qu'il n'existe pas de représentation normalisée des dictionnaires dans la mesure ou chaque analyseur, ne fonctionne qu'avec un dictionnaire dont la structure est dépendante des choix théoriques et conceptuels de cet analyseur et du degré de finesse que l'on veut atteindre dans les traitements.*
Mots clés: *Traduction automatique- dictionnaire électronique- conception de traduction automatique.*
Abstract: *This article examines, based on concrete examples, the problem of machine translation and the design of electronic dictionaries. The study showed that there is no standardized representation of dictionaries insofar as each analyzer only works with a dictionary whose structure is dependent on the theoretical and conceptual choices of this analyzer and the degree of finesse that we want to achieve in the treatments.*
Keywords: *Automatic translation- electronic dictionary- conception of automated translation.*

1. Introduction

En Traduction Automatique, qu'il s'agisse d'analyser un texte existant ou d'en fabriquer un, la machine utilise des dictionnaires et des lexiques. Toutefois, l'organisation de ces derniers n'est pas facile à concevoir.

2. Représentation du Mexique dans le dictionnaire

Une première approche consiste à voir dans les dictionnaires de simples listes de lexiques où chaque élément doit être illustré dans toutes les formes permises par la langue en question ; formes du singulier, du pluriel, du masculin, du féminin pour les substantifs ; les formes conjuguées des verbes dans tous les temps et tous les modes, toutes les formes adjectivales et adverbiales dérivées, etc. La classification et la représentation des mots simples repose sur les notions de parties du discours : verbes, noms, adjectifs, adverbes, prépositions, conjonctions, etc.

Pour obtenir un dictionnaire exhaustif conçu de cette façon nous pouvons imaginer facilement la taille qu'il aura et les coûts qui en découlent. En fait, les dictionnaires à usage humain, indispensables à la construction des dictionnaires électroniques, ne sont pas non plus complets. D'après le lexicographe Jean Dubois, dans la langue française, rien qu'en ce qui concerne les adjectifs en '-able', entre 1000 et 2000 n'ont jamais été répertoriées [cf. René Carré et al., 1991, p.143].

Les lacunes des dictionnaires se trouvent dans toutes les langues. Les lexicographes évitent généralement de les encombrer de mots et de formes qu'ils jugent inutiles dont le sens est évident. En fait, il est vrai que l'utilisateur humain est capable de formuler les mots qui manquent et de les interpréter. Mais pour la machine le phénomène est différent. Elle ignore tout simplement tout ce qui manque dans son lexique et se bloque par conséquent lors d'une consultation ou d'une génération.

Une seconde approche conçue par les informaticiens à ce propos consiste à ne présenter dans le dictionnaire que la forme canonique de chaque lexème de la langue et équiper les analyseurs de règles de dérivation pour obtenir les formes dérivées ; féminin, masculin, singulier, pluriel, conjugaison et adjectivalisation, etc.

En ce qui concerne l'adjectivalisation, il y a certaines correspondances régulières entre des classes de verbes et certaines formes adjectivales morphologiquement proches. Plusieurs sous-classes d'adjectifs sont ainsi répertoriées :

- Adjectivalisation verbale en 'eur' (AVeur) :

'Demander' ⟶ 'demandeur',

- Adjectivalisation verbale en 'ant' (AVant) :

'Enseigner' ⟶ 'enseignant'

- Adjectivalisation verbale en 'able' (AVable) :

'Accepter' ⟶ 'acceptable',

- Adjectivalisation verbale en ' ible' (AVible) :

'Convertir' ⟶ 'convertible'

- Adjectivaiisation verbale en 'uble' (AVuble) :

'Résoudre' ⟶ 'résoluble'

La dérivation des mots en '-tique' (comme : traitement automatique de l'information ⟶ informatique, automatisation du bureau ⟶ bureautique...).

En '-ation' (dériver ⟶ dérivation, différencier différentiation...),

En 'dé-' (régler ⟶ dérégler, connecter -> déconnecter...),

En 're-' (voir ⟶ revoir, lire relire...), etc…

Toutefois cette conception n'est pas saine non plus. Le traitement sur le radical est confronté aux différentes sortes de problèmes :

- La conjugaison des verbes irréguliers, ceux qui ne gardent pas le même radical (le schème) à tous les temps, comme le verbe 'aller' : 'vais', 'vas', 'allons', 'aille'., ou le verbe 'tenir' ou le radical sera :

- 'Tien' au présent de l'indicatif : 'je tiens', et de l'impératif : 'que je tienne'.
- 'Ten' au présent de l'indicatif : 'nous tenons'.
- 'Tiend' au futur simple : 'je tiendrai'.
- 'Tin' au passé simple : 'je tins'.

- Les singulier/pluriel et masculin/féminin irréguliers de certaines catégories de noms, exemples :

- Singulier/pluriel : une souris ⟶ des souris,

Un prix ⟶ des prix,

Un pneu ⟶ des pneus.

Du bétail ⟶ des bestiaux.

Un monsieur ⟶ des messieurs, etc.

- masculin féminin : un cerf ⟶ une biche.

Un cheval ⟶ une jument.

Un héros ⟶ une héroïne.

Un garçon ⟶ une fille,

Un taureau ⟶ une vache.

Un écrivain ⟶ une écrivain, etc...

- Les adjectifs : Il est difficile de définir une régularité de dérivation à partir des verbes pour une formalisation sereine. Par exemple, pour arriver du radical du verbe 'reproduire' aux adjectifs 'reproducteur' et 'reproductible 'on ne suit pas les mêmes règles pour dériver à partir du verbe 'convertir' les adjectifs 'convertisseur.' et 'convertible'.

Voir en Annexe 1 la représentation d'un échantillon du lexique du système DELAS : dictionnaire électronique développé au laboratoire d'automatique et linguistique de l'Université Paris VII

3. Le dictionnaire d'Oettinger (1959)

Oettinger est un mathématicien appliqué. Pour lui, c'est la construction des dictionnaires électroniques qui est prioritaire puisque le texte à traduire est une suite d'unités non structurées, dans le dispositif de traduction. L'analyse morphologique prime sur la syntaxe qui n'est à prendre en considération que secondairement. Oettinger choisit, pour des raisons linguistiques et techniques, de construire deux dictionnaires ; l'un de radicaux, l'autre de flexions : l'analyse morphologique permet l'identification des formes et son résultat est stocké pour être utilisé lors d'une analyse grammaticale ultérieure. Autrement dit, une approche hybride, mettant en œuvre les deux approches exposées précédemment.

La méthode préconisée par Oettinger consiste à réduire les formes à des paradigmes ; Un paradigme est formé par l'ensemble de toutes les formes fléchies d'un nom, d'un verbe ou d'un adjectif etc. Les membres d'un même paradigme sont identifiés par une racine commune définie comme la partie du mot et demeurant invariante sous flexion.

Cette organisation a été critiquée par certains auteurs d'autres systèmes de dictionnaire (notamment Lamb et al. 1961). Selon eux, la méthode d'Oettinger, qui sépare les suffixes avant le processus de consultation, présente un certain nombre d'inconvénients. La segmentation est souvent fausse, arbitraire, non fondée sur des principes structuraux. Elle n'est pas économique et implique d'utiliser davantage de racines, donc davantage d'entrées de dictionnaire.

- *Mots composés et expressions figées*

Pour traduire, l'humain repère les mots composés et les expressions figées pour pouvoir découper le texte en unités de traduction. Ces unités sont facilement repérables pour un individu qui a une bonne connaissance de la langue du texte qu'il traite. La définition que donne Tesnière au mot, semble vague et inadéquate dans ce contexte : *'. On ne saurait définir le mot par lui-même, mais seulement par les coupures qui en marquent le commencement et la fin'*. [Tesnière, 1976].

Martinet (1965) propose d'aller au-delà de 'l'écran du mot' pour analyser les unités minimales de signification. L'unité de signification (US) est la plus petite fraction signifiante. Elle peut comprendre un ou plusieurs mots. Autrement-dit, elle peut être un mot simple ou un mot composé.

Le mot composé est une expression obtenue par combinaison d'au moins deux mots simples et qui n'est pas nécessairement compositionnelle d'un point de vue syntaxique ou sémantique.

Il existe des mots composés appartenant à toutes les parties du discours :

- **Noms composés**

Fr : Pomme de terre ; coupe-papier ; compte courant…

En: Flesh and blood, fish and ships,

Ar: ساعي البريد, وقف إطلاق النار,.. [sa:çi Ibarird], [waqfu itlaïqi nna:ri]

- **Noms propres composés**

Fr : Côte d'ivoire, Sainte Marie de la mer, Bouches du Rhône,

En: New York, Sait Lake City,

Ar: عين صالح, برج باجي مختار,.. [çain salah], [burz ba:zi muxta:r|].

- **Adverbes composés (ou locutions adverbiales)**

Fr : en effet, de temps à autre, au moyen de.

En: a lot of, as a matter of tact...

Ar :...بيد أن [bajda ?anna)

- **Conjonctions composées**

Fr : parce que, tant et si bien que...

En: because of, as much...

Ar : ....عندما، حيثما [çindama:], [hajϴuma:]

- **Prépositions composées**

Fr : en face de, au-dessus de.

En : in front of,

Ar : ..قاب قوسين, في نا بعد, ما قبل التاريخ، [qa:ba qawsain], [fi: mabaçd], [ma: qabla tta:ri:x]

- **Adjectifs composés** :

Fr : bien portant, fidèle au poste.

En: blue eyed, ill mannered,

Ar:....أحمر قاتم, أصفر بني, متناهي الدقة, غير قابل للقهر [?ahmar qa.tim], [?aSfar bunni:], [mutana:hi ddiqa] ...

- **Verbes composés (ou locutions verbales)**

Fr : perdre la boussole, prendre en charge, ...

En: to look over, to go through, to over do,

Ar : أعادة النظر, أعادة الهيكلة, أدخل في, أخرج من... [? iça:datu nnadari], [? içadatu lhaikalati], [?udxila fi].

- **Expressions figées**

Fr : 'Mieux vaut tard que jamais', 'Aides-toi le ciel t'aidera,

En: 'Better late than never', 'it's raining cats and dogs'. 'God

helps who helps himself,

Ar:.'رب صدقة خير من ألف ميعاد, ' إن الطيور على أشكالها تقع ' [rubba Sudfatin xairun min alfi mi:qa:d], [?inna ttujura çala ajkaliha taqaçu],..

Afin d'éviter des traductions (mot à mot), comme celles-ci, aussi bien pour anglais-français que pour l'anglais-arabe.

- En : ' Better late than never'
  Fr : 1'- 'meilleur tard que jamais'
- En: 'It's raining cats and dogs'
  Fr : 2'- 'il-est pleuvant chats et chiens'

Haste makes waste.	الأستعجال يجعل تبدير
Better late than never	التأخر أفضل من لاشيء
A friend in need is a friend indeed	الصديق وقت الضيق
Birds of a feather feather flock together	سرب من نفس النوع سوية

Pour ce faire la machine doit être équipée d'un mécanisme qui prend chaque mot de la phrase et vérifie s'il ne forme pas avec les mots avoisinants un mot composé, une forme figée ou une * unité de traduction'.

Le mécanisme qui réalise ce genre de tâches est appelé, dans les milieux informatiques : 'automate'. Toutefois, la conception d'un tel automate n'est pas facile : s'il se limite à ne prendre qu'un nombre réduit de mots en analyse, les expressions figées contenant un nombre supérieur de mots lui échappent. Par contre, s'il est conçu d'une façon à traiter un nombre élevé de mots, il réduit considérablement la vitesse d'exécution des programmes et par conséquent l'efficacité de la machine et met en cause son utilité.

Il n'existe pas de représentation normalisée des dictionnaires dans la mesure ou chaque analyseur, ne fonctionne qu'avec un dictionnaire dont la structure est

dépendante des choix théoriques et conceptuels de cet analyseur et du degré de finesse que l'on veut atteindre dans les traitements. Voir en Annexe I la représentation des mots composés dans le dictionnaire.

En 1990, le Laboratoire d'Automatique Documentaire et Linguistique de l'Université Paris Vil (LADL) avait déjà recensé un nombre important de mots composés dans la langue française qui ont été catégorisés comme suit :

N Adj (carte grise) : 48 000

N de N (pomme de terre) : 28 000

N à N (pompe à eau) : 2 500

Adj N (petit pois) : 1 500

Prép N (à-côté) : 500

V N (lave-linge) : 1 000

N N (Télé-couleur) : 2 300

N à V (machine à écrire) : 200.etc.

Et plus de 6 000 adverbes composés comme : en bref, à dire vrai, etc. (Voir René Carré et al., 1991, p. 145).

Références

[1] CARRE René, J.F. Dégrémont, M'. GROSS, J.M. Pierrel, Gérard SABAfî, « Le traitement de l'écrit » in « Langage humain et machine », Paris, Presse du CNRS, 1991.

[2] KADRI Youcef, Ahmed BEN YAMINA, « Un système d'analyse syntaxico-sémantique du langage arabe non voyellée », mémoire d'ingénieur en informatique, Université d'Oran, 1992.

[3] BIROCHEAU Gaëlle, « L'ambiguïté Nom/Verbe dûe à la morphologie anglaise : Traitement automatique », in Bulag, n°21, Coordonné par Sylviane CARDEY-GREENFIELI), Centre Lucien Tesnière, Université de Franche-Comté, 1996.

[4] GHEN1MI Abdelouati. (2000), « Traduction automatique, français-anglais-arabe : problèmes et perspectives », mémoire de magistère. Université d'Oran, avril 2000.

[5] LAMB S.M., JACOBSEN W.H. (1961), «A High-speed Large-capacity Dictionary System», Journal ofMT6: 76-107.

[6] MARTINET A., « La linguistique synchronique », Paris, PUF, 1965.