

# **Le projet Longman-Liège : développements théssauriques**

par

T. HAYWARD  
A. MICHIELS  
J. NOEL

*Université de Liège - BELGIQUE*

473





En soumettant la totalité du dictionnaire à une telle procédure, on transformera le dictionnaire en un thésaurus. C'est là une des lignes de force du travail d'Amsler, travail qu'il ne nous est pas possible d'analyser plus en détail ici.

On contrastera cette approche de la création d'un thésaurus avec celle qui vient de donner lieu à la publication du LONGMAN LEXICON OF CONTEMPORARY ENGLISH (LOLEX; Mc Arthur 1981). L'ouvrage de Mc Arthur ne traite malheureusement que d'un sous-ensemble du LDOCE, puisque l'auteur s'est limité à 15.000 entrées contre les quelque 60.000 du LDOCE. Par ailleurs, bien que le LOLEX soit un sous-produit du LDOCE (en gros, les définitions en sont reprises), Mc Arthur n'a manifestement pas voulu ou pu exploiter la version informatique de ce dictionnaire. Bien au contraire il s'est donné, en se basant sur son intuition de lexicographe, des grilles conceptuelles et des catégories purement empiriques (ex : A : Life and Living Things, - Birds, A 73 : Waterbirds; L : Space and Time, - Geography, L 87 : Small Areas of Water).

Bornons-nous ici à déceler trois types d'insuffisances dans le résultat de la démarche de Mc Arthur, pour envisager ensuite une approche qui éviterait les faiblesses en question.

Quand on compare LDOCE et LOLEX, on est d'abord frappé par le faible taux de différenciation sémantique dans ce dernier. Il suffit de calculer la proportion terme/catégorie dans l'index du thésaurus pour en avoir une idée globale. De très nombreuses entrées de l'index du LOLEX ne renvoient qu'à une seule catégorie thésaurique. Par contre, les entrées du LDOCE sont largement polysémiques et bien sûr la reconnaissance des homographies contribue aussi à la multiplicité des sens repris dans le dictionnaire.

Un exemple frappant du manque de différenciation sémantique dans le LOLEX est l'amalgame malheureux des sens propre et figuré. On n'en compte pas moins de trois dans la catégorie

B 40 : the genitals and related organs

PRICK 1. the *penis* 2. a stupid *person*

BALLS 1. the *testicles* 2. *nonsense*

CUNT 1. the *vagina* or *vulva* 2. a *woman* or *girl* when regarded as a sexual object  
3. a stupid or bad *person*

(les noyaux sont soulignés).

Il est clair qu'une procédure tout à fait mécanique comme celle définie par Amsler aurait permis d'éviter cet écueil et de distinguer clairement l'emploi propre de l'emploi figuré.

L'examen des définitions des entrées appartenant à une même sous-classe thésaurique dans le LOLEX révèle souvent la présence d'un même noyau. Par exemple, dans la catégorie C161 : servants and menials, neuf items sur onze ont SERVANT comme noyau de définition. Les deux exceptions sont :

- 1.- SERVANT lui-même - ce qui est tout à fait normal puisqu'il s'agit d'un terme hiérarchiquement supérieur, qui détermine l'appartenance des autres termes à sa sous-classe.
- 2.- ATTENDANT, 1, dont le noyau est un quasi-pronom ("a person"), et où l'élément de signification repris par le noyau dans la définition des autres termes ("servant") est ici donné dans la *differentia* ("who . . . serves . . .").

D'autre part on peut se poser la question de savoir si la formulation de la deuxième définition d'ATTENDANT ("a person employed to look after and help visitors to a public place"), d'où l'élément de signification SERVE/SERVANT est absent, n'est pas le signe d'un classement erroné - la parenté avec des termes tels qu'EMPLOYEE ou CLERK, J.229 et J.234 dans LOLEX ("person employed . . .") étant peut-être plus pertinente. Quoi qu'il en soit, une procédure de création de thésaurus basée sur le repérage et la filiation des noyaux, permettrait une double vérification : de l'établissement des catégories thésauriques d'une part et de la formulation des définitions dans le dictionnaire-source d'autre part (concrètement, soit ATTENDANT, 2 n'est pas à sa place, soit il est inexactement défini).

On notera aussi que les liens hiérarchiques entre les sous-classes, qui peuvent être établis à l'aide de la procédure définie par Amsler, sont inexistantes dans le LOLEX. Par exemple, le noyau SERVANT permet d'entrer dans la sous-classe C163 : Slaves par le biais de la définition du mot SLAVE lui-même, définition dont le noyau est SERVANT ("slave : a *servant* who . . ."). Par la technique de filiation des noyaux, C163 est identifiable comme sous-classe de C161 : servants and menials. Par contre, dans LOLEX, les deux catégories sont au même niveau et la relation qui les unit n'apparaît pas.

Enfin, la technique de filiation des noyaux permet de repérer les intrus dans les sous-classes du LOLEX. Dans la catégorie B100 : having much fat, on trouve dix lexèmes, dont un, FLABBY, n'a pas le terme FAT dans sa définition. Par contre, dans B105 : heavily built, on trouve le terme DUMPY, qui est défini comme "short and FAT". Encore une fois, la conclusion est la même : soit ces deux termes sont mal définis, soit les catégories du LOLEX ne sont pas suffisamment homogènes.

Qu'on nous comprenne bien : nous ne prétendons nullement que l'annotation des définitions du LDOCE nous permettra de produire automatiquement un thésaurus impeccable. De toute évidence, la qualité d'un sous-produit thésaurique est tributaire de celle des définitions du dictionnaire-source. Ces dernières sont après tout produites intuitivement par des êtres humains et il n'y a aucune garantie que les définitions de deux termes sémantiquement proches feront apparaître cette parenté dans leur formulation. Toutefois, l'expérience d'Amsler sur le MPD et nos propres recherches sur le LOLEX

indiquent que les définitions sont un bon point de départ pour l'établissement de catégories thésaurologiques. D'autre part, quand on travaille sur l'ensemble du lexique d'une langue, l'utilisation de techniques informatiques devient impérative, non seulement pour établir toutes les relations pertinentes mais aussi pour en vérifier la cohérence. Le principal intérêt de la production d'un thésaurus à l'aide de moyens informatiques est qu'elle est riche en enseignements pour la rédaction des définitions, surtout à partir du moment où le dictionnaire-source s'astreint à un vocabulaire définitoire contrôlé, restriction qui, inévitablement, par le biais du mécanisme de la paraphrase, augmente la contribution de la syntaxe à l'énoncé des relations sémantiques.

#### REFERENCES

- LDOCE = LONGMAN DICTIONARY OF CONTEMPORARY ENGLISH (ed. : P. Procter, 1978).  
LOLEX = LONGMAN LEXICON OF CONTEMPORARY ENGLISH (T. Mc Arthur, 1981).  
MPD = MERRIAM-WEBSTER POCKET DICTIONARY.  
Amsler 1980 = Amsler, R. A., THE STRUCTURE OF THE MERRIAM-WEBSTER POCKET DICTIONARY, TR-164, University of Texas at Austin, décembre 1980.  
Michiels 1981 = Michiels, A. EXPLOITING A LARGE DICTIONARY DATA BASE, thèse de doctorat, Université de Liège, année académique 1981-1982.  
Olney et al. 1967 = Olney, J. et al., TOWARD THE DEVELOPMENT OF COMPUTATIONAL AIDS FOR OBTAINING A FORMAL SEMANTIC DESCRIPTION OF ENGLISH, SP-2766, System Development Corporation, Santa Monica, Cal., 1967.  
Olney 1968 = Olney, J., TO : ALL INTERESTED IN THE MERRIAM-WEBSTER TRANSCRIPTS AND DATA DERIVED FROM THEM, L-13579, System Development Corporation, Santa Monica, Cal., 1968.

NOTE

Facsimile de l'entrée LEAD<sup>1</sup>

**lead**<sup>1</sup> /li:d/ *v* **led** /led/ **1** [T1 (ALONG); IØ (ON)] to show (somebody) the way; go before or with (somebody) **2** [V3;T1 (to, into)] to guide in opinion or action; persuade; influence: *What led you to believe I was ill?* | *Her careless spending led her into debt* **3** [X9;L9] to be the means of (something) reaching a place, going through an area, etc.: *This path leads to the village* **4** [X9] to take or bring with, or as if with, force: *The horses were led into the yard* **5** [T1;IØ] to direct, control, or govern (an army, a movement, a meeting, etc.) **6** [T1;IØ] to be ahead in sports or games **7** [T1;IØ (with)] to start or open in (a match or game, esp. of cards, a dance, etc.) **8** [T1] to start or open a game, esp. of cards, with: *She led her highest card* **9** [T1] to experience or cause to experience (a kind of life): *He led a hard life* --see also LEAD ASTRAY, LEAD OFF, LEAD ON, LEAD TO, LEAD UP TO

**USAGE** To **lead** is to show someone the way by going first: *to lead an army* | *You lead and we'll follow*. To **guide** is to show the way and explain things: *to guide a party round the Taj Mahal*. To **drive** is either to control a moving vehicle, or to make animals move forward by going behind them: *driving the cattle to market*. To **steer** is to control the direction of a moving vehicle or boat. --see RIDE<sup>1</sup> (USAGE)