

ARTICLE 3: MORPHOSEMANTIQUE POUR L' APPARIEMENT DE TERMES DANS LE VOCABULAIRE MÉDICALE: APPROCHE MULTILINGUE

Cet article parle de la morphosémantique pour l'appariement de termes dans le vocabulaire médicale. Mais cet article s'intéresse plus particulièrement à la manière dont la morphosémantique peut contribuer à l'appariement multilingue des variantes terminologiques entre termes.

En effet, ils utilisent trois dispositifs:

- Un analyseur morphosémantique monolingue;
- Une table multilingue;
- 4 règles indépendantes de la langue.

Ce travail a été fait à partir d'un lexique spécialisé médical d'environ 29000 lexèmes.

Cet article est composé de 6 parties avec des sous-parties.

1- Variations terminologiques et relations lexicales entre termes

L'objectif de ce point est de montrer comment la morphologie et la terminologie peuvent collaborer pour optimiser l'appariement terminologie en corpus.

La Morphologie constructionnelle au service de la terminologie biomédicale:

- Elle apporte une aide incontournable dans l'association des termes reliés morphologiquement et établit les liens entre termes

Exemple:

- bactéries (adj) et bactéries (nom)
- hépatique (adj) et foie (nom) mais dans cet exemple, le lien n'est pas formellement identifiable

- Elle aide à faire une définition automatique des termes lorsqu'un terme est représenté par un lexème morphologiquement construit, il en existe une définition qui le relie à sa base.

L'intérêt de cette fonction est de deviner automatiquement le sens des mots inconnus quand ils sont conformes à une règle morphologique.

Exemple:

- bactéries: ce qui est relatif aux bactéries;
- anti-bactéries: contre les bactéries;
- hépatique: relatif au foie;
- antihépatique: contre l'hépatite

- Elle aide au regroupement des termes d'un corpus reliés lexicalement par synonymie, hyponymie ou approximation.

Exemple:

- abdominoscopie = laparoscopie
- thénatogène = mortifère
- entéralgie < abdominalgie
- cancéiforme ~ carcinoïde

Ces relations lexicales intéressent tout particulièrement les lexicologues, sémanticiens, logiciens et philosophes.

- Elle aide à la construction des termes d'un corpus en familles lexicales

Exemple:

- blastomycose (nom) synonyme of blastomycosique (adj)
- phlébodynie (nom) synonyme of phlébalgie (nom)

La morphologie des lexèmes spécialisés obéit à des règles constitutionnelles extrêmement proches dans toutes les langues européennes. (Lcobini, 2003)

La démarche multilingue s'applique au calcul des liens lexicaux entre mots composés savants du vocabulaire médical.

Ainsi nous constatons 3 types de ressources qui interagissent:

- un analyseur morphologique;
- une table établissant des relations de base entre les racines gréco-latines pouvant entrer dans la formation de mots;
- un système de règles calculant les relations lexicales entre termes.

La conception d'un analyseur doit être unique pour chaque nouvelle langue car la table est une donnée unique multilingue et que le système de règles est indépendant de la langue choisie. Cette approche a été implantée en français au moyen de DeriF dont quelques aspects viennent d'être dévoilés; sur un lexique d'environ 29 000

termes, elle donne lieu à l'émergence d'environ 3 000 familles lexicales.

2- Genèse

Les principes théoriques en morphologie lexicale permettent de déduire la définition d'un mot morphologiquement complexe en fonction de celui ou des ses constituants.

Quelques soit la langue européenne considérée, les mots complexes en biomédecine contiennent dans leur grande majorité des racines gréco-latines (gastr-; -phage; - hydr-).

Un EF partage sa catégorie et son sens avec l'entrée lexicale contemporaine auquel il supplée :

- gastr- signifie estomac en Fr et son type catégoriel est nom.

La terminologie médicale est aussi classée par les systèmes internationaux comme SNOMED, CIM-10, Mesh au moyen de relations lexicales synonymie, méronymie, (co)hyponymie...

3- Démarche

3 types de données et techniques qui répondent aux observations faites en 2:

- un ensemble réduit de règles générales et réunis dans une table et l'identification de ces Efs requiert l'intervention d'un analyseur morphologique.

- Analyseur morphologique monolingue

Le processus de décomposition d'un lexème complexe en constituant est une tâche monolingue dévolue à un analyseur morphologique qui peut fonctionner selon des approches diverses.

Exemples: hépatique (adj) = en relation avec le foie

- Ce tableau illustrent différentes situations et différentes langues dans lesquelles les Efs servent à construire des lexèmes spécialisés suffixés, préfixés ou composés.

Lang.	Affixatio n/ Compos ition ⁹	<i>traducti on</i>	EFs abstrait
IT	epatico	<i>hépatiq</i>	HEPAT
FR	buccal	<i>ue</i>	BUCC
EN	analges(ic)	<i>buccal analgési</i>	ALGES THERM
DE	Hypothe rmie	<i>que hypothé</i>	CEPHAL
ES	intracef al(ico)	<i>rmie intracép halique</i>	
(1)	IT	gastro	<i>gastre</i> GASTR
(2)	FR	ectom	<i>ctomi</i> ,
(3)	EN	ia	<i>e</i> ECTO
(4)	DE	bucco	<i>bucco</i> MI
(5)	ES	dent(a	<i>dentai</i> BUCC,

ire)	<i>re</i>	[DENT
therm	<i>therm</i>]10
oalgés	<i>oalgé</i>	THER
(ia)	<i>sie</i>	M,
Therm	<i>therm</i>	ALGES
otaxis	<i>otaxie</i>	THER
braqui	<i>brach</i>	M,
cefalo	<i>ycéph</i>	TAXI
	<i>ale</i>	BRAC
		HY,CE
		PHAL

4- Résultats pour le français

Les résultats ont été obtenus à partir d'un lexique totalisant 29 noms, adjectifs, et adverbes du vocabulaire spécialisé, collectés à partir de diverses sources librement accessibles en ligne ou mises en la disposition des projets UMLF et Vumef. l'analyseur morphosémantique DériF est le réalisateur de la chaîne de traitement en français.

L'analyse morphologique d'un lexème construit sur une base elle-même construite est donc hiérarchisée.

Le résultat est un triplet, la première partie retrace sous forme

crochetée l'historique des étapes d'analyse, la seconde réunit les lexèmes résultats obtenus à chaque étape, et la troisième est constituée d'une formulation en langue naturelle de la relation morphologique liant l'input à son (ses) constituant(s)immédiat(s).

Exemple: relations lexicales candidates pour gastralgie
Nom

gastralgie/NOM => [[gastr N*] [algie N*] NOM]

(gastralgie/NOM, algie/N*)

" douleur (du -- liée au) estomac "

Constituants = /gastr/algie/

Type = maladie

Relations possibles = (eql:gastr/algo, eql:gastr/algés,
eql:gastr/odyn, eql:stomac/algie, eql:stomac/algo,
eql:stomac/algés, eql:stomac/odyn, eql:stomach/algie,
eql:stomach/algo, eql:stomach/algés, eql:stomach/odyn,
isa:abdomin/algie, isa:abdomin/algo, isa:abdomin/algés,
isa:abdomin/odyn, see:entéro/algie, see:entéro/algo,
see:entéro/algés, see:entéro/odyn, see:gastr/ite, see:gastr/
ose, see:hépat/algie, see:hépat/algo, see:stomach/ite,
see:hépat/algés, see:hépat/odyn, see:pancréat/algie,
see:pancréat/algo, see:pancréat/algés, see:pancréat/odyn,
see:stomac/ite, see:stomac/ose, see:stomach/ose).

La recherche d'information dans le domaine de la biomédecine a déjà fait l'objet d'expérimentation par (Schulz et al., 1999) et (Hanh et al., 2001) qui se servent également des Efs qu'ils appellent 'subwords'. Leur analyse morphologique est réduite à un simple découpage linéaire qui ne permet pas d'associer une définition à l'input. Cette approche a des inconvénients majeur, qui est celui de tout système basé sur l'utilisation de contraintes linguistiques et demandant la gestion des exceptions.

Cette approche nécessite une validation humaine à trois niveaux au moins:

-vérifier la pertinence des analyses, pour valider les pseudo-définitions et surtout pour contrôler la pertinence des relations lexicales de base dans la table des Efs. Éviter des annotations trop spécifiques sur des Efs polyréférentiels en médecine.

Améliorer cette approche en créant de nouvelles règles, de nouvelles relations lexicales de base et un nouveau module.