

# Terminologie médicale en français : Les projets UMLF et VUMeF

Pierre Zweigenbaum

LIMSI, CNRS, Orsay, France

Journée CISMéF  
Rouen, 4 février 2008

- 1 Contexte
  - Enjeux
  - État des lieux
- 2 Besoins dans le contexte français
  - Existence : couverture
  - Lisibilité : libellés en français
  - Facilité d'usage : lien avec la langue
- 3 Projets français
  - Langue : UMLF (2002–2005)
  - Terminologie : VUMeF (2003–2006)
  - La suite : InterSTIS (2008–2010)

# Terminologie médicale en français : enjeux

- Terminologie médicale et systèmes d'information
  - Terminologie, textes, information : recherche, indexation
  - Recherche, navigation dans et entre les terminologies
  - Quelles terminologies utiles  
(dépend des tâches et des utilisateurs) ?
- Terminologies de langue française
  - Couvrir les terminologies en usage en France (resp. Canada, Suisse, Belgique...)
  - Gérer la langue française
  - Multilinguisme

# Terminologie médicale en français : enjeux

- Terminologie médicale et systèmes d'information
  - Terminologie, textes, information : recherche, indexation
  - Recherche, navigation dans et entre les terminologies
  - Quelles terminologies utiles  
(dépend des tâches et des utilisateurs) ?
- Terminologies de langue française
  - Couvrir les terminologies en usage en France (resp. Canada, Suisse, Belgique...)
  - Gérer la langue française
  - Multilinguisme

# État des lieux : le projet UMLS

## Objectif

Faciliter la recherche et l'intégration d'informations provenant des multiples sources d'information biomédicales électroniques

- Bibliothèque nationale de médecine américaine (NLM, Bethesda)
- Projet démarré en 1986, première version en 1989
- Ressource distribuée gratuitement (mais respect des droits)
- <http://www.nlm.nih.gov/research/umls/>

# UMLS : Unified Medical Language System

## Metathesaurus : *terminologies*

- union raisonnée d'une centaine de terminologies biomédicales

## Réseau sémantique : *types sémantiques, relations sémantiques*

- structuration unifiante superposée à ces terminologies

## Lexique Specialist : *langue*

- informations morphosyntaxiques sur les mots biomédicaux
- outils d'analyse de la langue (anglaise)

# Le Metathesaurus de l'UMLS

- Inclut et relie une centaine de terminologies biomédicales
- 1,5 million de concepts (CUI)
- 5,5 millions de termes différents (LUI)
- 3,6 millions de libellés(SUI) en anglais, 126 000 en français

*(UMLS 2007AC)*

- Un must dans tous les travaux de (trans)codage
- Permet d'explorer de nombreuses terminologies
  - Thésaurus MeSH (12 langues : anglais, français, ...)
  - Classification CIM-10 (anglais, allemand, néerlandais)
  - Nomenclature SNOMED CT (anglais, espagnol)
  - ...

# Systemes propriétaires

- Voir le programme de l'après-midi (collaborations commerciales)
  - Couverture des terminologies françaises ?
  - Prise en charge du français ?

- 1 Contexte
  - Enjeux
  - État des lieux
- 2 Besoins dans le contexte français
  - Existence : couverture
  - Lisibilité : libellés en français
  - Facilité d'usage : lien avec la langue
- 3 Projets français
  - Langue : UMLF (2002–2005)
  - Terminologie : VUMeF (2003–2006)
  - La suite : InterSTIS (2008–2010)

# Existence : couvrir les terminologies en usage en France

- Un serveur de terminologie doit intégrer les terminologies en usage en France (resp. Canada, Suisse, Belgique...)
- L'UMLS ne le fait pas pour toutes
  - Exemple : la Classification commune des actes médicaux (CCAM) n'est pas dans le Metathesaurus de l'UMLS

↪ *Lier à l'UMLS les terminologies françaises qui en sont absentes*

# Lisibilité : fournir des libellés en français

Les versions françaises...

- ne sont pas toujours **incluses** dans l'UMLS
  - ICD-10 → CIM-10 en français
  - SNOMED International (v3.5) → version française
  - NANDA → ANADI en français
- ne sont pas toujours aussi **complètes** que les originales
  - MeSH anglais : 23616 descripteurs et termes vedettes,  
27804 synonymes en 2005
  - MeSH français : 12185 synonymes en 2004\*
- n'**existent** pas toujours
  - SNOMED CT

↔ *Accroître la quantité et la qualité des termes français dans les terminologies internationales (VUMeF)*

# Facilité d'usage : lien avec la langue

- Recherche de termes dans une terminologie
  - *infarctus du myocarde*  
→ I21 Infarctus aigu du myocarde, I21.0...
- Recherche de termes approchants
  - *infarctus myocardique* → *infarctus du myocarde*
  - *crise cardiaque* → *infarctus du myocarde*
- Aide à l'indexation
  - texte → termes repérés dans le texte
  - indexation MeSH, codage CIM-10 ou SNOMED

↪ *Fournir des ressources lexicales sur le français médical (UMLF)*

- 1 Contexte
  - Enjeux
  - État des lieux
  
- 2 Besoins dans le contexte français
  - Existence : couverture
  - Lisibilité : libellés en français
  - Facilité d'usage : lien avec la langue
  
- 3 Projets français
  - Langue : UMLF (2002–2005)
  - Terminologie : VUMeF (2003–2006)
  - La suite : InterSTIS (2008–2010)

# Langue: UMLF (2002–2005)

- Projet ACI 2002–2005
- [STIM/AP-HP](#) (coordinateur)
- U Rouen, U Rennes, U Nancy/CNRS
- Société Vidal
- HU Genève

*(Zweigenbaum et al. 2005)*

# UMLF : mots du lexique médical

- Lexique : mots en français
  - forme fléchiée → forme canonique
    - *abdominaux* → *abdominal*
  - mot dérivé → mot de base
    - *chitineux* → *chitine*
  - mot composé → décomposition structurée

## adactylie/N

- 1 [ a+ [ dactyl N\* ] +ie N ]
- 2 affection caractérisée par l'absence de doigt

# Terminologies francophones et codage :

## Le projet VUMeF

- 2003–2006, Vidal, U Rouen, 12 partenaires (RNTS)
- A participé à l'extension du versant francophone de terminologies internationales (MeSH, SNOMED)
- Méthodes de codage MeSH (Rouen, Marseille)
- Méthodes de repérage de variantes de termes (CNRS, Inserm)

*(Darmoni et al. 2003)*

# VUMeF : Partenaires

- Société [Vidal](#) (coordinateur), [CHU Rouen](#) (resp. scientifique)
- CHU Rennes, CHU Marseille, CH Mulhouse
- U Nancy/CNRS, CNRS Toulouse, Inserm DISC, Inserm U729
- Société Synomia
- HU Genève, SFINM Sherbrooke

# VUMeF : Extension des versions françaises de MeSH et SNOMED

- Traduction française du MeSH par le DISC de l'Inserm (depuis 1986)
- Pendant le projet VUMeF, avec l'aide de l'INIST :
  - 12185 synonymes en 2004
  - 30832 synonymes en 2006
  - Passage à une forme accentuée, casse mixte, termes plus naturels
    - *ULCERE CORNEE* → *Ulcère de la cornée; Ulcère cornéen*
- Traduction française de la SNOMED Internationale par le SFINM (Sherbrooke) avec l'aide d'A Buemi (CH Mulhouse ; AP-HP)
  - soutien du projet VUMeF à la traduction

# VUMeF : Aide au codage MeSH

- CHU Rouen, CHU Marseille
- Repérage de termes MeSH dans des textes
- Choix des mots-clés majeurs / mineurs

# VUMeF : Repérage de variantes de termes français

- Variantes  $N (Prep) N_1 \leftrightarrow N A_1$  de termes MeSH
- 309 variantes nouvelles proposées, précision  $\sim 54 \%$

# occurrences				$N(Prep) N_1$			$NA_1$	
$n_t$	$a_t$	$n_m$	$a_m$					
1	121	1	0	cancer		bronche	cancer	bronchique
1	44	1	0	évaluation		technologie	évaluation	technologique
1	19	1	0	tumeur		thymus	tumeur	thymique
1	2	1	0	médecine		reproduction	médecine	reproductif
1	1	1	0	cancer		thorax	cancer	thoracique
1	1	1	0	professionnel	en	éducation	professionnel	éducatif
1	1	0	1	besoin	en matière de	nutrition	besoin	nutritif
1	1	0	1	chambre	de	malade	chambre	malade
1	1	0	1	diarrhée	de	bovin	diarrhée	bovin
1	1	0	1	fragment	de	peptide	fragment	peptidique
1	1	0	1	infection	dans	oeil	infection	ophtalmique
1	1	0	1	médicament	pour	animal	médicament	animal
1	1	0	1	toxine	de	choléra	toxine	cholérique
1	2	0	2	cellule	de	estomac	cellule	gastrique
1	3	0	3	résistance	de	champignon	résistance	fongique

# VUMeF : recherche de traductions françaises de termes anglais

Termes MeSH en anglais + corpus parallèle Santé Canada =

MeSH américain	MeSH français	corpus parallèle Santé Canada
cheek bones	Apophyse zygomatique	pommettes des joues
Death	Mort	Décès
Asian	Groupe d'ascendance asiatique	Asiatique
Atomic Energy	Énergie nucléaire	énergie atomique
vegetables	Plantes potagères	légumes
fertility rate	Taux natalité	taux de fécondité
Infant Mortality	Mortalité du nourrisson	Mortalité infantile
dizziness	Sensation vertigineuse	sensation de vertige
vibration	Vibration	vibratoire
vulnerable populations	Populations vulnérables	populations à risque
West Nile virus	Virus du Nil occidental	VNO
wounds	Plaies et traumatismes	blessures
construction materials	Matériaux construction	matériaux de construction
birth weight	Poids naissance	Poids à la naissance

# La suite...

Serveur de terminologies : InterSTIS (2008–2010)