

L'ANNOTATEUR IMEDIATE

Un outil pour l'encodage des dossiers médicaux en SNOMED-CT

Damien De Meyere

Assistant de recherche

Institut Langage et Communication



Séminaire du CENTAL

28 avril 2017

CONTEXTUALISATION



L'information collectée au cours du **parcours clinique** d'un patient est aujourd'hui enregistrée au **format électronique**.



Mais près de 70% de l'information consiste en du **texte libre, difficilement exploitable** par un programme informatique.



Au vu des **multiples sources** disponibles, il est **fastidieux d'aggréger** les informations à propos d'un patient ou groupe de patients.



Ces informations pourraient être **partagées** et réutilisées pour des **objectifs multiples** : suivi du patient, essais cliniques, détection d'effets indésirables, qualité des soins, etc.

- Dans les hôpitaux : **structuration de l'information médicale** en recourant à des **classifications** du type ICD.
- Des efforts ont été faits pour tenter d'**automatiser** le codage des documents dans le cadre du **Résumé clinique minimum**.
- L'**ICD** et **ICPC** sont des **classifications** élaborées à des fins administratives et d'épidémiologie.
- Ces classifications ne sont pas les meilleures ressources pour l'indexation sémantique (**manque de précision et/ou de couverture**).

SNOMED-CT

- **350 000 concepts** cliniques (vs 13 000 pour l'ICD-9, 69 000 pour l'ICD-10), répartis en **19 axes** pour décrire les symptômes, les circonstances, les maladies, les diagnostics, les interventions chirurgicales, les résultats et les décisions thérapeutiques, etc.
- Possibilité de combiner plusieurs codes pour référer à un concept complexe (**postcoordination**).

Ex. : "Venous insufficiency great saphenous vein both legs"

→ SNOMED-CT: 234050002|Venous insufficiency of leg
|:363698007|Finding site|=181402008|Great saphenous vein|,
272741003|Laterality|=51440002|Bilateral|

→ ICD-10-CM: I87.2 *Venous insufficiency (chronic) (peripheral)*

→ SNOMED-CT est plus approprié pour l'**indexation** des données patients.

- Utilisé dans plus de **50 pays** → réelle interopérabilité des données médicales.
- Existe en **plusieurs langues**, mais uniquement sous **forme parcellaire pour le français**, notamment au Canada et en Belgique.
- En Belgique, SNOMED-CT fait partie de la stratégie d'interopérabilité dans le cadre du Plan national **e-Santé 2013-2018** → adoption (très) progressive, volume de données peu important.

LE PROJET IMEDIATE

Interoperability of medical data through information extraction and term encoding

innoviris.brussels 
empowering research



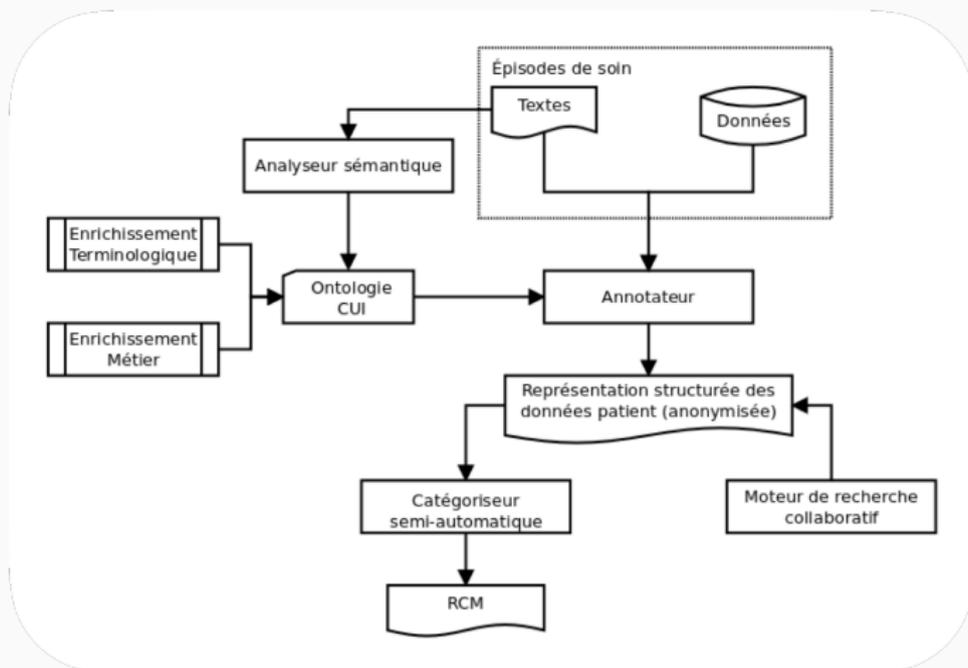
Hôpital
Erasme



abrumet



LE PROJET IMEDIATE EN UN COUP D'OEIL



La plupart des études se concentrent sur l'ICD, le MeSH ou l'UMLS.

Pour SNOMED-CT, quelques références existent pour l'anglais...

[Patrick et al., 2007, Jung et al., 2009, Lee et al., 2010, Khare et al., 2012, Kate, 2013]

... mais relativement peu pour le français.

[Pereira et al., 2006, Pereira et al., 2009]

· Challenges :

- Pas de ressources d'entraînement suffisantes pour cette tâche
- Collecter un corpus demande beaucoup d'efforts :
 - diversité de codes, sous-langages et types de textes ;
 - confidentialité des données médicales ;
 - accord inter-annotateurs.
- Les ressources terminologiques en français sont moins étendues que pour l'anglais
- Haut degré de variation dans les textes médicaux
 - Maladie de Crohn, Crohn, Entérite de Crohn, Entérite granulomateuse, Entérite régionale
 - Thrombose artérielle cérébelleuse, Thrombose de l'artère cérébelleaire
 - Méso-sigmoïde, Sigmoïde et son méso

TRADUCTION DE SNOMED-CT EN FRANÇAIS

- **Collaboration tripartite**
 - terminologie (TERMISTI/ULB)
 - sciences de la santé (ESP/ULB)
 - traitement automatique du langage (CENTAL/UCL)
- **Différents dictionnaires :**
 - **SNOMED-CT** (subset belge FR) : 44 000 entrées
 - **Ressources existantes** : ICD-9-CM, ICD-10-CM, Erasme, ICPC-2 (mappings)
 - **Termisti**
 - **CENTAL**

Dépôt centralisé des ressources terminologiques (UCL/PRISME)

The screenshot displays a software interface for managing terminological resources. At the top, there are tabs for different terminologies: SNOMED-CT, ICD-9-CM, ICD-10-CM/PCS, ICPC2, ERASME, TERMISTI, and UMLS. The main area shows a hierarchical list of terms. Two terms are highlighted with red boxes and arrows:

- Synonymes:** Points to a red box around the following terms:
 - 1235 anneau gastrique ajustable
 - 1236 AGA
 - 1237 anneau de gastroplastie modifiable
 - 1238 AGM
- Lien vers SNOMED:** Points to a red box around the following term:
 - SNOMED-CT 24883002 bypass gastrique

Other visible terms in the list include: 1228 rectosigmoidoscopie de dépistage, 1229 tumeur de l'angle duodénojéjunal, 1230 tumeur de l'angle duodénojéjunal, 1231 ulcères jéjunaux multiples, 1232 ulcérations duodénojéjunales, 1239 gastroentérologie, 1240 gastrectomie en manchon, 1241 gastrectomie en manchette, 1242 sleeve gastrectomie, 1247 bypass en Y, 1248 superobésité, 1249 superobésité, 1250 superobésité, 1251 adhérences multiples sous-mésocoliques, 1252 difficultés d'ascension de l'intestin grêle, 1253 adiposité abdominale, 1254 taux de ghréline, 1256 hormone de la faim, UMLS 7349 Ghéline, 1257 transit oesogastroduodénal, 1258 TOGD, 1259 TOGD postopératoire, 1260 transit baryté oesogastroduodénal, 1261 chirurgie bariatrique, 1262 chirurgie de l'obésité, 1266 mini-bypass gastrique, 1267 bypass en Oméga.

· Méthodologie semi-automatique

- **Collecte** automatique de termes (CENTAL)
- **Nettoyage** manuel des entrées (TERMISTI)
 - Survie après splénectomie → splénectomie
 - Pylori chez les patients → pylori
- **Évaluation et catégorisation** par 3 binomes d'experts (ESP)
 - Résection biliaire et hépatique → résection biliaire + résection hépatique
- **Suggestion** de codes SNOMED-CT (CENTAL)
- **Raccrochage** à SNOMED-CT par 3 binomes d'experts (ESP)

· Spécialités

- **Gastroentérologie** : collecte automatique de termes dans un corpus de textes médicaux, à l'aide de la méthode de [Watrin and François, 2011]
 - **Cooccurrences de structures nominales** constituées de **formants** typiques de la gastroentérologie : *jéjun*, *oesoph*, *gastr*, *stomac*, *cholangi*, *bilie*, *biliair*, *hépat*, *foie*, etc.
- **Chirurgie de l'obésité** : océrisation d'un index en français
- **Médecine générale** : océrisation d'un index anglais-français

Source	Nb termes	Codage automatique	Codage manuel	Catégorisé
Index gastroentérologie	908	405	503	non
Corpus documentaire I	5 117	508	4 609	oui
Corpus documentaire II (chirurgie obésité)	83	10	73	non
Index médecine générale	7 870	7 870	0	non
TOTAL TERMISTI	13 978 (100%)	8 793 (63%)	5 185 (37%)	

Table: Constitution du dictionnaire Termisti

Termes de gastroentérologie : 6 108 (44 %)

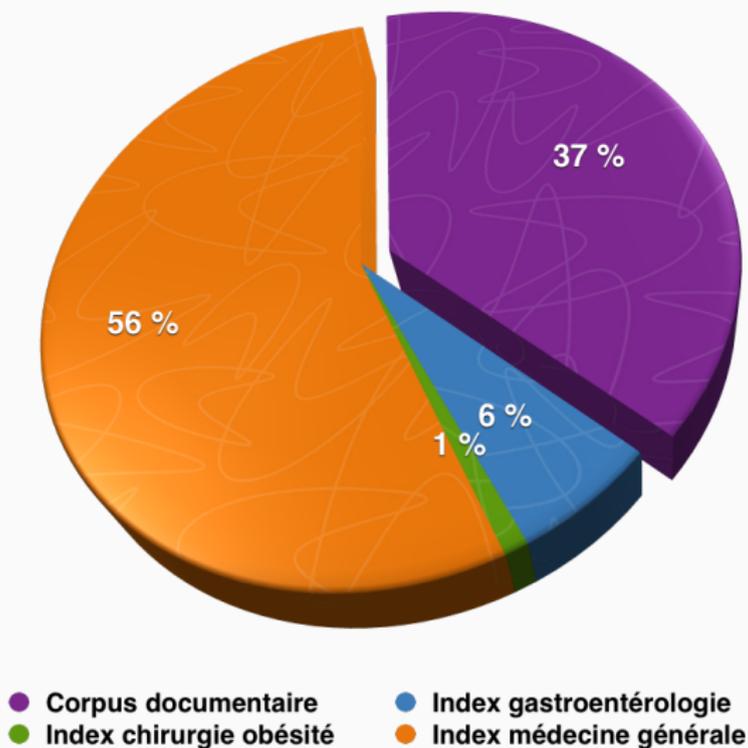


Figure: Constitution du dictionnaire Termisti

Lemaire, N., François, T., Debongnie, J.-C., De Meyere, D., Fauquert, B., Klein, T., Fairon, T. Van Campenhoudt, M. (2015). **L'enrichissement terminologique d'usage du projet iMediate : une collaboration tripartite terminologie / TAL / sciences de la santé.** In Actes du Second Congrès international du Réseau de Lexicographie (5-7 octobre 2015), Universidade de Santiago de Compostela. Ibéroamericana.

- **Méthodologie majoritairement automatique**
 - **Extraction** des termes en français enregistrés dans l'UMLS (ICPC, MedDRA, MeSH, MTHMST et WHO-ART)
 - **Mapping** des identifiants de concepts UMLS (CUI) vers les concepts SNOMED-CT
- **Toutes spécialités**

Source	Nb termes	Technique
UMLS FR	78 424	Mappings CUI UMLS
Snomed 3.5 VF	78 851	Mappings CUI UMLS
Qualifiers < ESP	1 272	Flexion manuelle (Termisti)
Qualifiers < ICD-10	997	Traductions EN-FR + flexion manuelle (Termisti)
TOTAL CENTAL	159 544	

Table: Constitution du dictionnaire Cental

PLUS-VALUE DES DIFFÉRENTES STRATÉGIES D'ENRICHISSEMENT

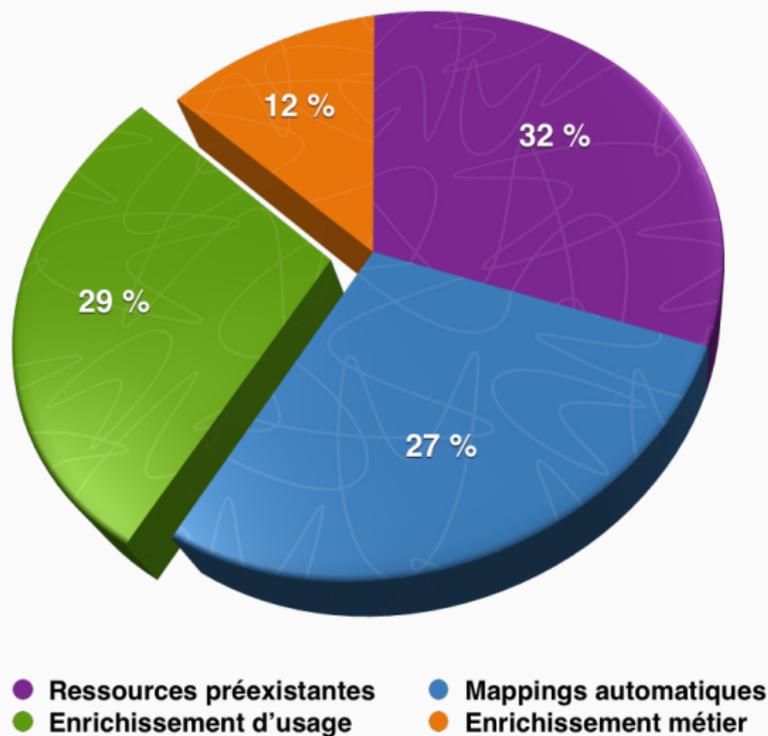


Figure: Vue par source pour les termes en bilio-, bilie- et biliair-

ANNOTATION EN SNOMED-CT

Un système de codage en SNOMED-CT qui applique
une base terminologique de façon flexible.

- Algorithme basé sur le contenu d'OntoEditor
- Prise en compte de la variation linguistique inhérente à la pratique des équipes de soins

- **Suppression d'expressions génériques** typiques des classifications
 - Douleur, *non classée ailleurs*
 - Intoxication alimentaire *SAI*
- **Conversion des diacritiques en lettres non-accentuées.**
- **Suppression des mots vides**
 - Quelques exceptions : vitamine A, *DU* perforé
- **Conversion en bas de casse**
 - Quelques exceptions : résection *du* sigmoïde vs *DU* perforé
- **Généralisation de variantes récurrentes :**
 - Chiffres romains ↔ chiffres arabes ;
 - Mots composés écrits avec un espace, un tiret ou agglutinés (*cardio-myopathie, cardiomyopathie, etc.*)

- **Importation** des entrées nettoyées dans l'annotateur :
 - Pour chaque terme dans la ressource terminologique :
 - **Tokenisation** (découpe en mots)
 - **Stemming** (rationalisation par suppression récursive de suffixes : oesophagienne/oesophage → oesophag)

- Dépouillement manuel des **séries synonymiques déjà enregistrées** dans OntoEditor
- Sur cette base, génération **dynamique** de **variantes**
- Dérivations adjectivales : 228 séries, 657 synonymes
 - oedeme du *muscle*/*myo*-oedeme/oedeme *musculaire*
 - douleurs aux *gencives*/douleur *gingivale*
 - abcès *cholécystique*/abcès de la *vésicule biliaire*
 - débit *sanguin hépatique*/débit *hém*-hépatique
 - acidité *gastrique*/acidité de l'*estomac*/acidité *stomacale*
 - mucus *rhino-pharyngé*/mucus du *nez* et du *pharynx*
 - circulation *splanchnique*/circulation des *viscères*
 - colonisation de la *peau*/colonisation *cutanée*

- Graphes Unitex

- Négation

- *Pas de* signe de décompensation cardiaque
 - Le test de latences multiples du sommeil *ne montre pas de* somnolence diurne et *exclut* une narcolepsie

- Antécédents personnels

- Accident de moto à 16 ans
 - Dépression nerveuse (*il y a 10 ans*)
 - Hypertension artérielle *traitée*

- Antécédents familiaux

- Néoplasme du colon *chez une tante*
 - *Son père* souffrirait d'asthme et un oncle paternel de diabète
 - *Mère* alcoolique

- Modèle CRF (C. Hansart)

- Segmentation en rubriques

- Environ **2 400 protocoles** et **1 600 lettres de sortie** annotés manuellement
 - Schéma d'annotation en **dix rubriques** (affections actuelles, antécédents personnels, résultats d'analyses, examens, interventions, traitement de sortie, conclusion, etc.)
 - **Traits** (position de la phrase dans le texte, marqueurs de temps, ponctuation, comparaison de vecteurs de mots, lexique utilisé (verbes spécifiques à certaines rubriques, médicaments, etc.)).

choc anaphylactique après piqûre d'abeille

1 2 3 4 5 6

DICTIONNAIRE

choc,39579001+PROB+2
 choc,241931004+PROB+4
 choc,68031002+PROB+2
 allerg,39579001+PROB+2
 allerg,241931004+PROB+4
 react,39579001+PROB+2
 anaphylact ,39579001+PROB+2
 anaphylact ,241931004+PROB+4
 piqur,241931004+PROB+4
 systol,68031002+PROB+2
 abeill 241931004+PROB+4
 ...

BUFFER



choc anaphylactique après piqûre d'abeille

1 2 3 4 5 6

DICTIONNAIRE

choc,39579001+PROB+2
 choc,241931004+PROB+4
 choc,68031002+PROB+2
 allerg,39579001+PROB+2
 allerg,241931004+PROB+4
 react,39579001+PROB+2
 anaphylact ,39579001+PROB+2
 anaphylact ,241931004+PROB+4
 piqur,241931004+PROB+4
 systol,68031002+PROB+2
 abeill 241931004+PROB+4
 ...

BUFFER

39579001+PROB+2
 241931004+PROB+4
 67195008+PROB+2



choc **anaphylactique** après piqûre d'abeille

1 2 3 4 5 6

DICTIONNAIRE

choc,39579001+PROB+2
choc,241931004+PROB+4
choc,68031002+PROB+2
allerg,39579001+PROB+2
allerg,241931004+PROB+4
react,39579001+PROB+2
anaphylact ,39579001+PROB+2
anaphylact ,241931004+PROB+4
piqur,241931004+PROB+4
systol,68031002+PROB+2
abeil 241931004+PROB+4
...

BUFFER

39579001+PROB+2
241931004+PROB+4
67195008+PROB+2



choc anaphylactique après piqûre d'abeille

1 2 3 4 5 6

DICTIONNAIRE

choc,39579001+PROB+2
choc,241931004+PROB+4
choc,68031002+PROB+2
allerg,39579001+PROB+2
allerg,241931004+PROB+4
react,39579001+PROB+2
anaphylact ,39579001+PROB+2
anaphylact ,241931004+PROB+4
piqur,241931004+PROB+4
systol,68031002+PROB+2
abeil 241931004+PROB+4
...

BUFFER

39579001+PROB+2
241931004+PROB+4



choc anaphylactique après **piqûre** d'abeille

1 2 3 4 5 6

DICTIONNAIRE

choc,39579001+PROB+2
choc,241931004+PROB+4
choc,68031002+PROB+2
allerg,39579001+PROB+2
allerg,241931004+PROB+4
react,39579001+PROB+2
anaphylact ,39579001+PROB+2
anaphylact ,241931004+PROB+4
piqur,241931004+PROB+4
systol,68031002+PROB+2
abeil 241931004+PROB+4
...

BUFFER

241931004+PROB+4



choc anaphylactique après piqûre d'abeille

1 2 3 4 5 6

DICTIONNAIRE

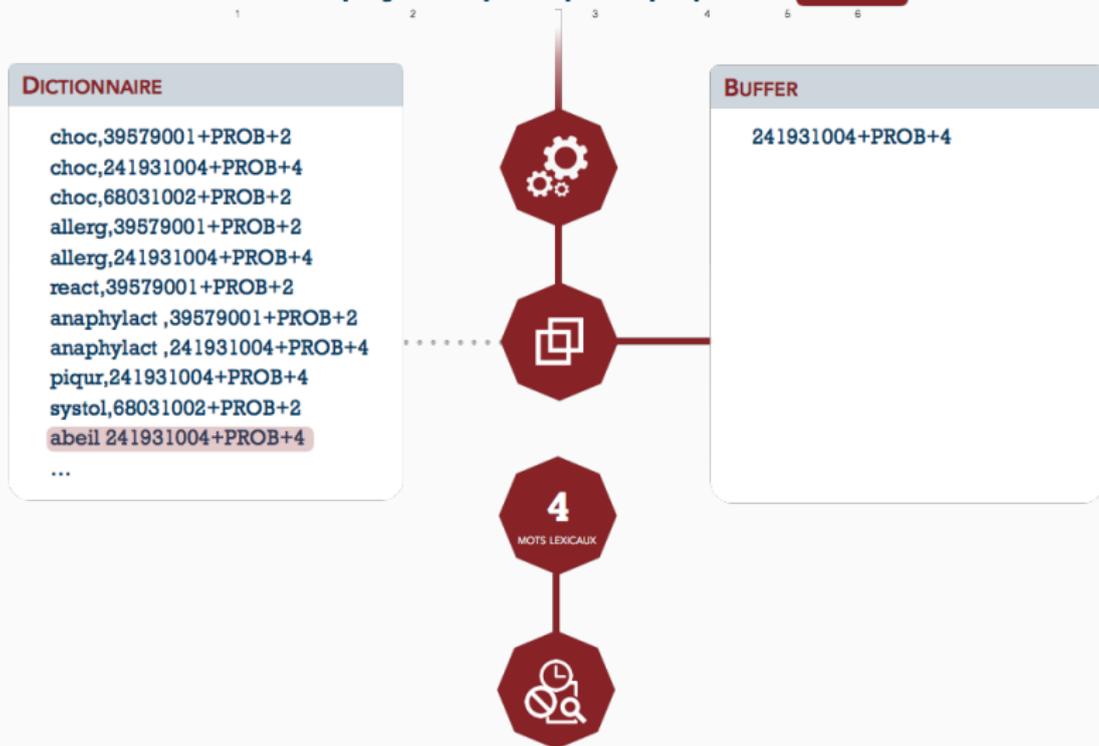
choc,39579001+PROB+2
choc,241931004+PROB+4
choc,68031002+PROB+2
allerg,39579001+PROB+2
allerg,241931004+PROB+4
react,39579001+PROB+2
anaphylact ,39579001+PROB+2
anaphylact ,241931004+PROB+4
piqur,241931004+PROB+4
systol,68031002+PROB+2
abeil 241931004+PROB+4
...

BUFFER

241931004+PROB+4



choc anaphylactique après piqûre d'abeille



choc anaphylactique après piqûre d'abeille

1 2 3 4 5 6

DICTIONNAIRE

choc,39579001+PROB+2
choc,241931004+PROB+4
choc,68031002+PROB+2
allerg,39579001+PROB+2
allerg,241931004+PROB+4
react,39579001+PROB+2
anaphylact ,39579001+PROB+2
anaphylact ,241931004+PROB+4
piqur,241931004+PROB+4
systol,68031002+PROB+2
abeil 241931004+PROB+4
...

BUFFER

241931004+PROB+4



The screenshot shows a web browser window with the title "Annotateur iMediate". The browser's address bar is empty. The page has a dark blue header with the text "iMediate Annotateur" and a refresh icon on the right. The main content area is white and features the heading "Annotateur". Below the heading, there are two input options: "Taper un texte" and "Télécharger un fichier". A large text area is provided for input, with the instruction "Tapez le texte à annoter :". To the right, under the heading "Paramètres", there are three sections: "Types d'entités à extraire" with buttons for "Diagnostic", "Procédure", "Anatomie", "Organisme", "Qualifier", "Autre", and "Non catégorisé"; "Dictionnaires à utiliser"; and "Segmentation en rubriques". At the bottom right of the parameter section, there are three buttons: "Reset", "Annoter", and "Résumé structuré (JSON)". At the bottom left of the page, the footer text reads "Webmaster : CENTAL, Université catholique de Louvain".

Annotateur

Taper un texte Télécharger un fichier

Tapez le texte à annoter :

Paramètres

Types d'entités à extraire

Diagnostic Procédure Anatomie Organisme

Qualifier Autre Non catégorisé

Dictionnaires à utiliser

Segmentation en rubriques

Reset Annoter

Résumé structuré (JSON)

Webmaster : CENTAL, Université catholique de Louvain

The screenshot displays the iMediate interface. On the left, a text editor window shows a medical note with several terms highlighted in green. On the right, a structured view of the text is shown, with entities categorized into Introduction, Antécédents, and Examens. Each entity is represented by a checkmark, a label, and a corresponding code and description.

Cher Confrère, J'ai revu votre patiente à ma **consultation** ce [DATE] pour une visite de contrôle à un mois d'une **résection sigmoïdienne selon Hartmann** et une **colostomie temporaire**. Veuillez trouver ci-joint un compte-rendu: Antécédents médicaux **Polyarthrite - Plastie mammaire** en [DATE]. De façon générale, pas de **complications postopératoires**. Un **bon état** général est signalé hormis une rapide **asthénie** et quelques **vertiges en station debout**; la **température corporelle est normale**. L'**appétit augmente** progressivement mais **persistance de nausées** (prise de Primpéran le matin). Un **amaigrissement** de 8 kilos est signalé. L'évolution apparaît très favorable et les soins locaux identiques sont à poursuivre. Un **test de microbiologie** à effectuer dans 3 semaines est prescrite ce jour. J'ai programmé sous réserve de votre accord une **hospitalisation** le [DATE] pour bilan, et si tout va bien comme je le pense, une intervention le [DATE] pour **remise en continuité après Hartmann**.
Je ne manquerai pas de vous tenir informé aussitôt, Bien confraternellement

Entités reconnues **16**

Introduction

- ✓ **Consultation** **[procedure]** 223475005 Consulting with (procedure)
- ✓ **Colostomie temporaire** **[procedure]** 11429006 Consultation
- ✓ **Résection sigmoïdienne selon Hartmann** **[TEMP]** 1733@TERMISTI
- ✓ **Colostomie temporaire** **[procedure]** 386829005 Temporary colostomy (procedure)

Antécédents

- ✓ **Polyarthrite** **[disorder]** 417373000 Inflammatory polyarthropathy
- ✓ **Polyarthrite** **[disorder]** 416956002 Polyarthrits
- ✓ **Plastie mammaire** **[procedure]** 33496007 Reconstruction of breast

Examens

- ✗ **Complications postopératoires** **[disorder]** 385486001 Postoperative complication
- ✓ **Bon état** **[finding]** 102499006 Fit
- ✓ **Asthénie** **[TEMP]** 27159@ERASME
- ✓ **Asthénie** **[finding]** 13791008 Lassitude
- ✓ **Vertiges en station debout** **[finding]** 407645004 Dizziness on standing up (finding)
- ✓ **Température corporelle est normale** **[finding]** 87273009 Temperature normal
- ✓ **Appétit augmente** **[finding]** 72405004 Increased appetite (finding)

EXEMPLES DE SÉQUENCES EXTRAITES

Termes reconnus	Entrées du dictionnaire
cancer bronchique épidermoïde du lobe droit	carcinome épidermoïde bronchique du lobe moyen du poumon droit
varices oesophagiennes	varice de l'oesophage
oedème aigu pulmonaire	oedème pulmonaire aigu
la vidange de la vessie est incomplète	vidange incomplète de la vessie
méso-sigmoïde	sigmoïde et son méso
résection sigmoïdienne (Hartmann)	résection sigmoïdienne selon Hartmann
examen colonoscopique	colonoscopie
tumeurs du col de l'utérus malignes	Tumeur maligne du col de l'utérus

PREMIÈRE ÉVALUATION

- Extraction terminologique
- Validation et raccrochage des candidats-termes
- Développement de l'annotateur

→ Évaluation globale encore à mener

- Constat : l'annotation d'un corpus de référence demande des efforts considérables !
 - 5 lettres de sortie anonymisées (gastroentérologie)
 - 2 paires d'experts pour souligner les termes présents dans le texte→ Déjà 258 termes cliniques !
- Une évaluation binaire est trop restrictive (décomposition d'une entité complexe en plus petites unités, nature hiérarchique de SNOMED-CT, etc.) → évaluation plus fine.

Évaluation	# d'occurrences
Détection complète	179 (69.38%)
Exact match	103 (39.92%)
Post-coordination	61 (23.64%)
Moins spécifique	1 (0.39%)
Trop spécifique	2 (0.78%)
Incorrect	12 (4.65%)
Détection partielle	79 (30.62%)
Moins spécifique	30 (11.63%)
Incomplète ou erronée	49 (18.99%)
Total	258 (100%)

Évaluation experts	# d'occurrences
Détection complète	179 (69.38%)
Exact match	103 (39.92%)
Post-coordination	61 (23.64%)
Moins spécifique	1 (0.39%)
Trop spécifique	2 (0.78%)
Incorrect	12 (4.65%)
Détection partielle	79 (30.62%)
Moins spécifique	30 (11.63%)
Incomplète ou erronée	49 (18.99%)
Total	258 (100%)

Les ressources terminologiques existantes sont partiellement représentatives de la pratique langagière effective → haut degré de variation linguistique dans les rapports médicaux.

Évaluation experts	# d'occurrences
Détection complète	179 (69.38%)
Exact match	103 (39.92%)
Post-coordination	61 (23.64%)
Moins spécifique	1 (0.39%)
Trop spécifique	2 (0.78%)
Incorrect	12 (4.65%)
Détection partielle	79 (30.62%)
Moins spécifique	30 (11.63%)
Incomplète ou erronée	49 (18.99%)
Total	258 (100%)

- Dans presque 25% des cas, notre programme n'est pas parvenu à extraire la séquence complète (→ terme long), mais identifie avec succès tous les constituants

Ex. : *[Oedème] des [membres inférieurs]
[douleur] au niveau de la [partie supérieure de l'abdomen] en [ceinture]
[souscostale] avec [irradiation] dans le [flanc droit]*

- Beaucoup de termes médicaux ont une structure compositionnelle, constituée d'un concept noyau acceptant un ou plusieurs modificateurs.

→ Le mécanisme de la post-coordination SNOMED-CT pour limiter la taille des ressources terminologiques ?

Évaluation experts	# d'occurrences
Détection complète	179 (69.38%)
Exact match	103 (39.92%)
Post-coordination	61 (23.64%)
Moins spécifique	1 (0.39%)
Trop spécifique	2 (0.78%)
Incorrect	12 (4.65%)
Détection partielle	79 (30.62%)
Moins spécifique	30 (11.63%)
Incomplète ou erronée	49 (18.99%)
Total	258 (100%)

- Erreurs dans les ressources base de données terminologique (erreurs de mapping ou codes dépréciés)
- Caractère ambigu de certains termes :
 - Alcool → alcoolisme (*disorder*), boisson alcoolisée (*substance*) ou substance désinfectante (*substance*)
- Le système détecte le constituant principal, mais le code assigné n'est pas suffisamment précis
 - *Adénocarcinome* transitionnel de la vessie

- Corpus de 125 lettres de sortie.
- Extraction de ± 400 termes médicaux de plus de 3 mots avec l'approche souple.
- Comparaison avec une approche qui ne se base que sur le stemming.
- **Gain obtenu grâce à la reconnaissance souple : 54%.**

De Meyere, D. ; Klein, T. ; François, T. ; Debongnie, J.-C. ; Radulescu, C. ; Mbengo, N. ; Ouro Koura, M. ; Coppieters 't Wallant, Y. ; Fairon, C. (2015). **Automatic annotation of medical reports using SNOMED-CT: a flexible approach based on medical knowledge databases**. In: Proceedings of the 7th Language Technology Conference (Poznań, Poland, du 27/11/2015 au 29/11/2015), p. 519-523. <http://hdl.handle.net/2078.1/171882>

CONSIDÉRATIONS FINALES

Enrichissement terminologique

- Étude de la langue médicale dans son usage effectif.
- Limitation des ressources terminologiques actuelles.
- Puits sans fond ? Mais ressources de grande valeur.

· Annotateur

- Apport intéressant de la reconnaissance souple.
- Utilité de la post-coordination SNOMED-CT.
- Nombreux défis liés à la variation linguistique
- Substrat d'une approche statistique ?
- Exploitation dans un autre contexte ?

QUESTIONS ?

REFERENCES

- Jung, S., S. Kim, S. Yoo, and J. Choi, 2009.
Toward the Automatic Generation of the Entry Level CDA Documents.
Journal of Korean Society of Medical Informatics, 15(1):141–151.
- Kate, R. J., 2013.
Towards Converting Clinical Phrases into SNOMED CT Expressions.
Biomedical Informatics Insights, 6(Suppl. 1):29–37.
- Khare, R., Y. An, J. Li, I. Song, and X. Hu, 2012.
Exploiting semantic structure for mapping user-specified form terms to snomed ct concepts.
In *Proc. of the 2Nd ACM SIGHIT Int. Health Informatics Symposium, IHI '12*. New York, NY, USA: ACM.
- Lee, D., F. Lau, and H. Quan, 2010.
A method for encoding clinical datasets with SNOMED CT.
BMC Medical Informatics and Decision Making, 10(1):53–64.
- Patrick, J., Y. Wang, and P. Budd, 2007.
An Automated System for Conversion of Clinical Notes into SNOMED Clinical Terminology.
- Pereira, S., P. Massari, A. Buemi, B. Dahamna, E. Serrot, M. Joubert, and S. J. Darmoni, 2009.
F-MTI : outil d'indexation multi-terminologique : application à l'indexation automatique de la SNOMED Int.
In *Risques, Technologies de l'Information pour les Pratiques Médicales*, number 17 in Informatique et Santé. Springer Paris, pages 57–68.

- Pereira, S., A. Névéol, P. Massari, M. Joubert, and S. Darmoni, 2006.
Construction of a semi-automated ICD-10 coding help system to optimize medical and economic coding.
Studies in Health Technology and Informatics, 124:845–850.
- Watrin, P. and T. François, 2011.
N-gram frequency database reference to handle MWE extraction in NLP applications.
In Proc. of the 2011 ACL Workshop on MultiWord Expressions.