


Das Forum für ICT im Gesundheitswesen  
Le forum pour les TIC dans le système de santé

 @eHealthSummit

[www.ehealthsummit.ch](http://www.ehealthsummit.ch)



**SGMI SSIM SSMI**

Schweizerische Gesellschaft für Medizinische Informatik  
Société Suisse d'Informatique Médicale  
Società Svizzera d'Informatica Medica  
Swiss Society for Medical Informatics

**STADE DE SUISSE  
BERN**  
11.-12. SEPT. 2018

# Automatic annotation of French medical narratives with SNOMED-CT concepts

Christophe Gaudet-Blavignac, Mmed, BSc CS, UNIGE/HUG

 @chrisgaubla

In cooperation with



**ehealthsuisse**  
Koordinationen Plan & Rollen  
Organe de coordination Confédération-Canton  
Organi di coordinamento Confederazione-Cantoni

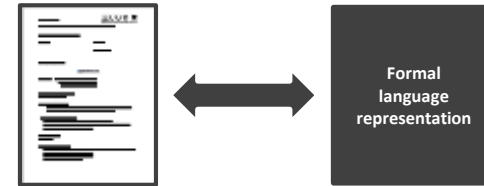
**IHE**  
SUISSE  
Integrating  
the Healthcare  
Enterprise

pharmaSuisse 

**VGIch**  
Vereniging Geneeskundige Informatica Vereniging

# INTRODUCTION

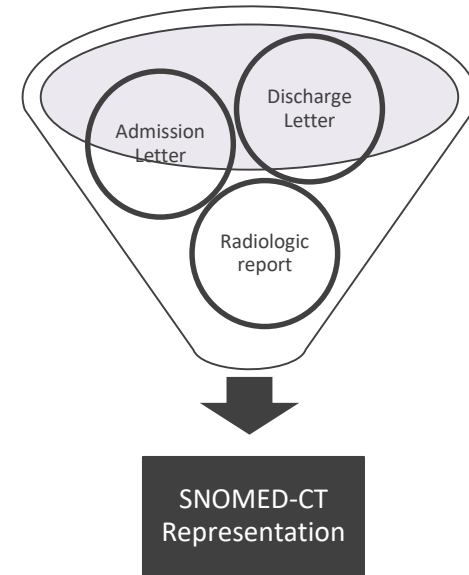
- Medical data is composed of both structured data and narrative data (free text)
- Narratives:
  - contain valuable information
  - rarely reused
  - constitute a challenge for interoperability



An automatic translation from narrative data to a formal language such as SNOMED-CT would provide better reusability and interoperability

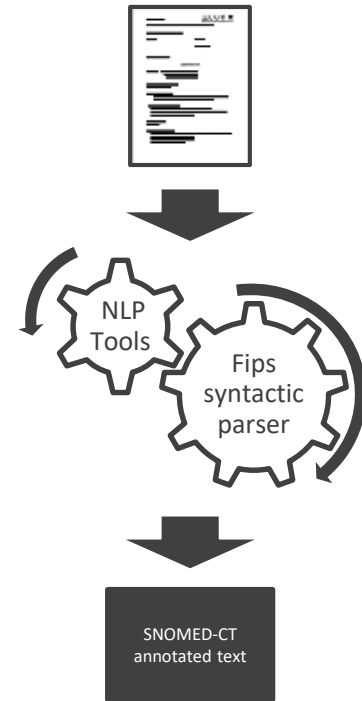
# INTRODUCTION

- First step toward automatic translation:  
**Annotation**
- Propose a method for automatic annotation of French medical narratives with SNOMED-CT concepts
- Preliminary evaluation



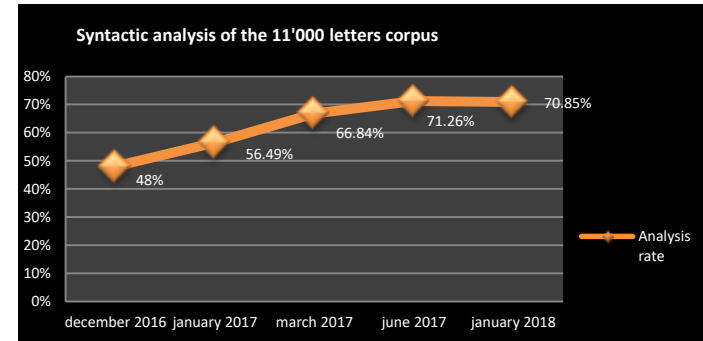
# METHOD

- Tool adaptation
  - Adapt existing tool (UNIGE) to:
    - Process medical French text
    - Output a SNOMED-CT annotated text
  - Steps needed
    - Creation of lexico-semantic resources
    - SNOMED-CT as a natural language
    - Create the annotation task



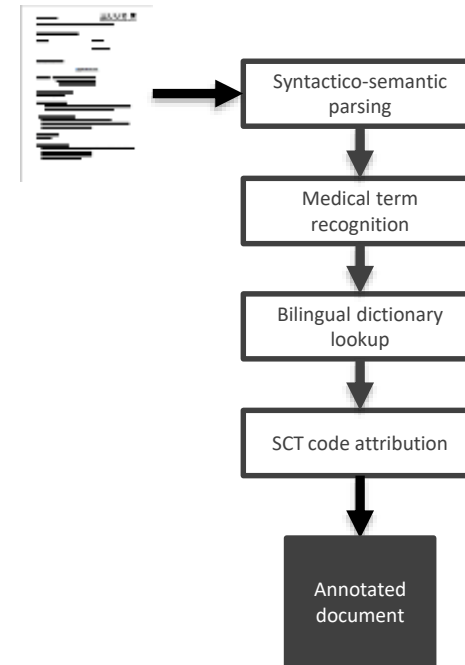
# METHOD

- Electronic dictionaries
  - 4'831 lexemes (*pneumonie*), 13'519 inflected forms
  - 5'659 collocations (*angine à streptocoque*)
  - 5'974 bilingual mappings French  $\leftrightarrow$  SNOMED-CT



# METHOD

- Automatic annotation
  - Parsing of initial text
  - Medical term recognition
  - Bilingual dictionary lookup
  - SNOMED-CT code attribution
- Corpus
  - Corpus of 11'000 French discharge summaries



# AUTOMATIC ANNOTATION

- Automatic annotation

Words	4'481'191
Annotated terms	892'787
Unique SNOMED-CT concepts	7'569
Annotated terms per sentence	4.17

# AUTOMATIC ANNOTATION

Un cathétérisme cardiaque droit sera réalisé chez le patient.

Le laboratoire révèle une protéine C réactive élevée.

Le patient prend des anti-inflammatoires, traitement qui est poursuivi durant l'hospitalisation.



SNOMED-CT code	French term
40403005	cathétérisme du cœur droit
398166005	réalisé
116154003	patient

SNOMED-CT code	French term
261904005	laboratoire
263857004	révélé
61425002	protéine c-réactive
260399008	élevé

SNOMED-CT code	French term
116154003	patient
373283003	anti-inflammatoire
266714009	poursuivre le traitement
32485007	hospitalisation



# PRELIMINARY EVALUATION

- **5** discharge letters (**1'820** words) manually annotated with SNOMED-CT codes
- Automatic annotation for **421** SNOMED-CT concepts
- Evaluation results (perfect match):
  - **Precision: 0.7173**
  - **Recall: 0.5171**
  - **F-score: 0.6009**
- **Limitations**
  - Size of the corpus
  - 1 annotation expert
  - Evaluation correct only if perfect match

# CURRENT WORK

- Post-coordination
  - Post-coordinate annotated terms using the SNOMED-CT compositional grammar
  - «The Compositional Grammar Specification and Guide specifies a standard syntax for representing individual clinical meanings using either pre-coordinated or post-coordinated expressions»

## Pre-coordinated

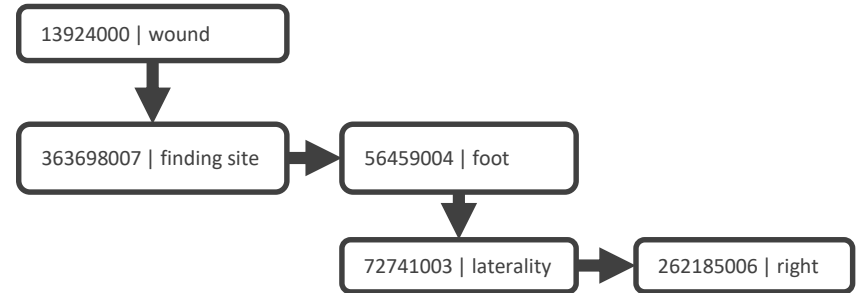
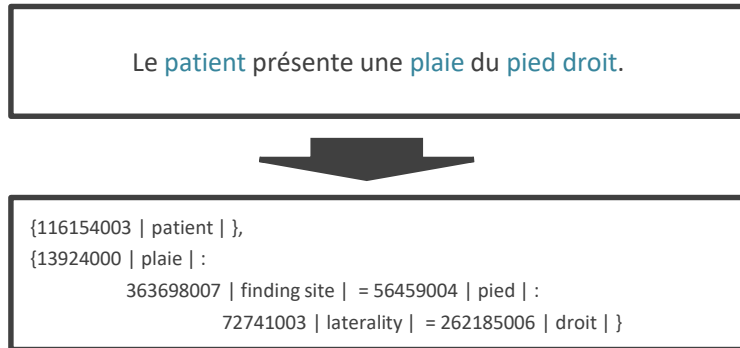
74400008 | Appendicitis (disorder)|

## Post-coordinated

```
=== 18526009 | Disorder of appendix (disorder) | :  
  {  
    363698007 | Finding site | =  
      66754008 | Appendix structure (body structure) | ,  
    116676008 | Associated morphology | =  
      23583003 | Inflammation (morphologic abnormality) |  
  }
```

# EXAMPLE

- After implementation of SNOMED-CT compositionnal grammar and some post-coordination rules



# CONCLUSION

- Automatic annotation of French medical narratives with SNOMED-CT concepts is a demanding but rewarding task
  - Usage of a specifically modified syntactico-semantic parser
  - Development of large lexico-semantic resources
  - Implementation of specific rules
  - Preliminary evaluation is encouraging
- Annotation is the first step toward automatic translation of narratives into post-coordinated SNOMED-CT sentences.

# REFERENCES

- [1] S. M. Meystre, C. Lovis, T. Bürkle, G. Tognola, A. Budrionis, and C. U. Lehmann, "Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress."
- [2] "Support : SNOMED International." [Online]. Available: <https://ihtsdo.freshdesk.com/support/home>. [Accessed: 20-Mar-2017].
- [3] J. Patrick, Y. Wang, and P. Budd, "An Automated System for Conversion of Clinical Notes into SNOMED Clinical Terminology."
- [4] P. Ruch, J. Gobeill, C. Lovis, and A. Geissbühler, "Automatic medical encoding with SNOMED categories.," BMC Med. Inform. Decis. Mak., vol. 8 Suppl 1, p. S6, 2008.
- [5] G. Zuccon et al., "Automatic Classification of Free-Text Radiology Reports to Identify Limb Fractures using Machine Learning and the SNOMED CT Ontology.," AMIA Jt. Summits Transl. Sci. Proc. AMIA Summit Transl. Sci., vol. 2013, pp. 300–4, Jan. 2013.
- [6] A. Nguyen, J. Moore, G. Zuccon, M. Lawley, and S. Colquist, "Classification of pathology reports for Cancer Registry notifications," Stud. Health Technol. Inform., vol. 178, no. May 2014, pp. 150–156, 2012.
- [7] M. Torii, K. Wagholikar, and H. Liu, "Using machine learning for concept extraction on clinical documents from multiple data sources.," J. Am. Med. Inform. Assoc. JAMIA, vol. 18, no. 5, pp. 580–587, 2011.
- [8] B. Koopman, G. Zuccon, A. Nguyen, A. Bergheim, and N. Grayson, "Automatic ICD-10 classification of cancers from free-text death certificates," Int. J. Med. Inf., vol. 84, pp. 956–965, 2015.
- [9] B. Riedl, N. Than, and M. Hogarth, "Using the UMLS and Simple Statistical Methods to Semantically Categorize Causes of Death on Death Certificates.," AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp., vol. 2010, pp. 677–81, 2010.
- [10] P. Ruch, J. Gobeill, C. Lovis, and A. Geissbühler, "BMC Medical Informatics and Decision Making Automatic medical encoding with SNOMED categories."
- [11] D. De Meyere et al., "Automatic annotation of medical reports using SNOMED-CT: a flexible approach based on medical knowledge databases," 2015.
- [12] A. R. Aronson, "Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program."
- [13] R. J. Kate, "Towards Converting Clinical Phrases into SNOMED CT Expressions.," Biomed. Inform. Insights, vol. 6, no. Suppl 1, pp. 29–37, 2013.
- [14] E. Wehrli and L. Nerima, "The fips multilingual parser," in Language Production, Cognition, and the Lexicon, Springer, 2015, pp. 473–490.
- [15] C. Hansart, D. De Meyere, P. Watrin, A. Bittar, and C. Fairon, "CENTAL at SemEval-2016 Task 12: a linguistically fed CRF model for medical and temporal information extraction," Proc. 10th Int. Workshop Semantic Eval. SemEval-2016, pp. 1286–1291, 2016.

# Thank you for your attention!

