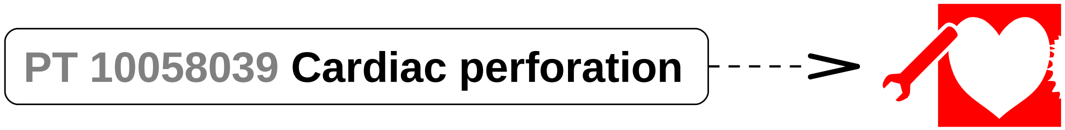
Combining semantic and lexical methods for mapping MedDRA to VCM icons

Jean-Baptiste Lamy
Rosy Tsopra



PT 10058039 **Cardiac perforation**

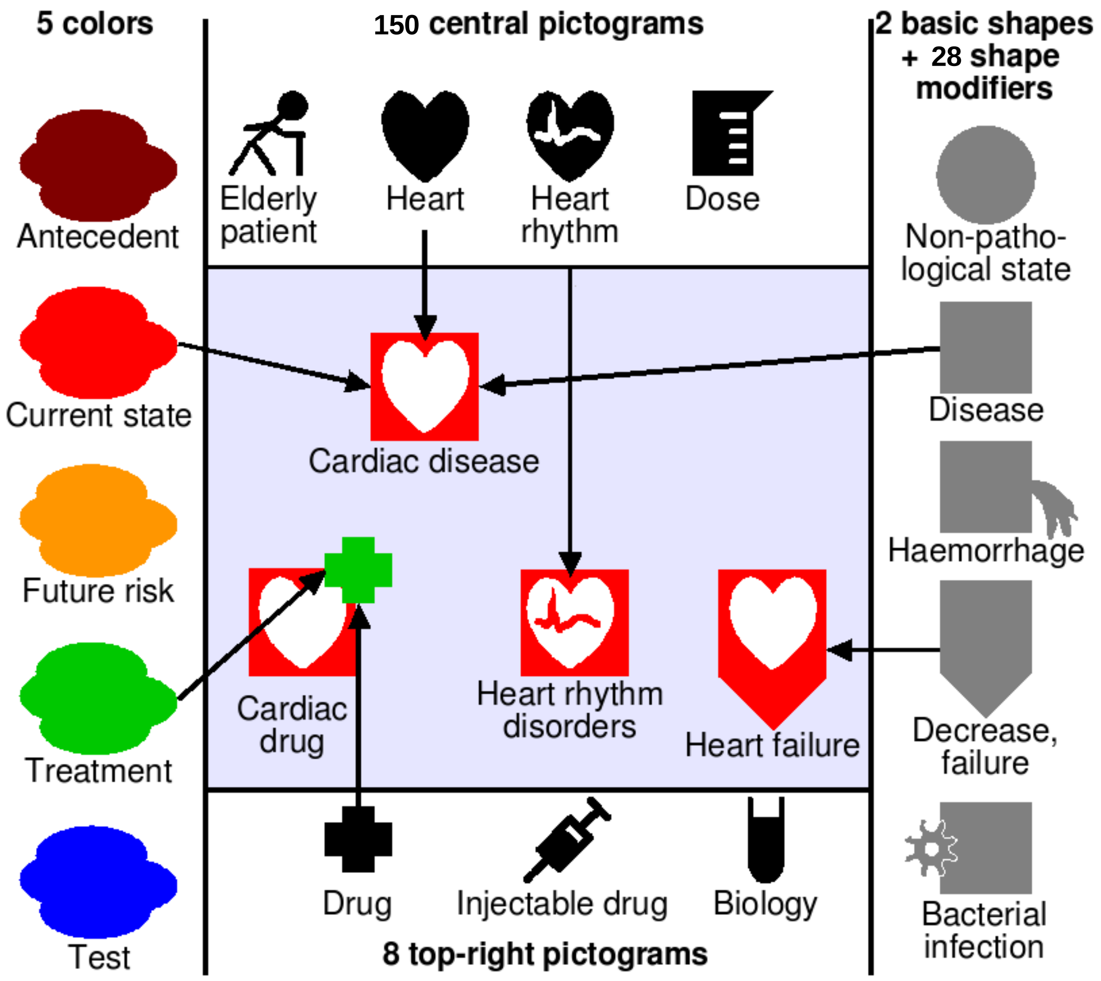
LIMICS
Université Paris 13, Sorbonne Paris Cité, 93017 Bobigny
Sorbonne Universités, Paris
INSERM UMRS 1142

Introduction

- **Medical terminologies**
  - Essential for semantic interoperability
  - But difficult for Humans!
- **=> we developed since 10 years VCM, an iconic language for representing medical concepts**
  - Not as precise as terminologies, but can be used for enriching text or illustrating terms
- **Requires mapping between icons and terminologies**
  - Semantic methods for terminologies with a formal semantics (*e.g.* SNOMED CT [MEDINFO])
  - Other terminologies requires more complex methods
  - Here, we will focus on MedDRA :
    - Used for coding adverse effects
    - Multiaxial classification without formal semantics
  - => lexico-semantic method

VCM
(Visualization of Concept in Medicine)

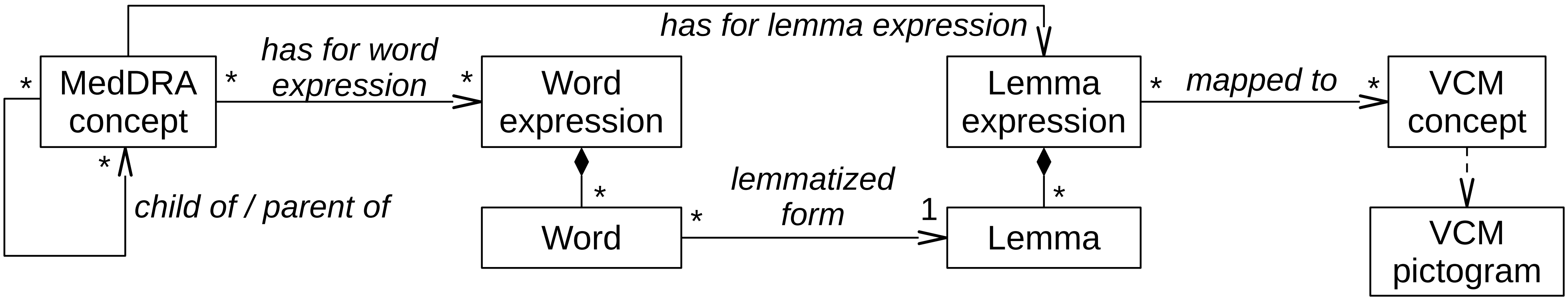
- **An iconic langage for medical concepts** [BMC]
  - Symptoms
  - Disorders
  - Treatments
  - Exams
  - Adverse effects



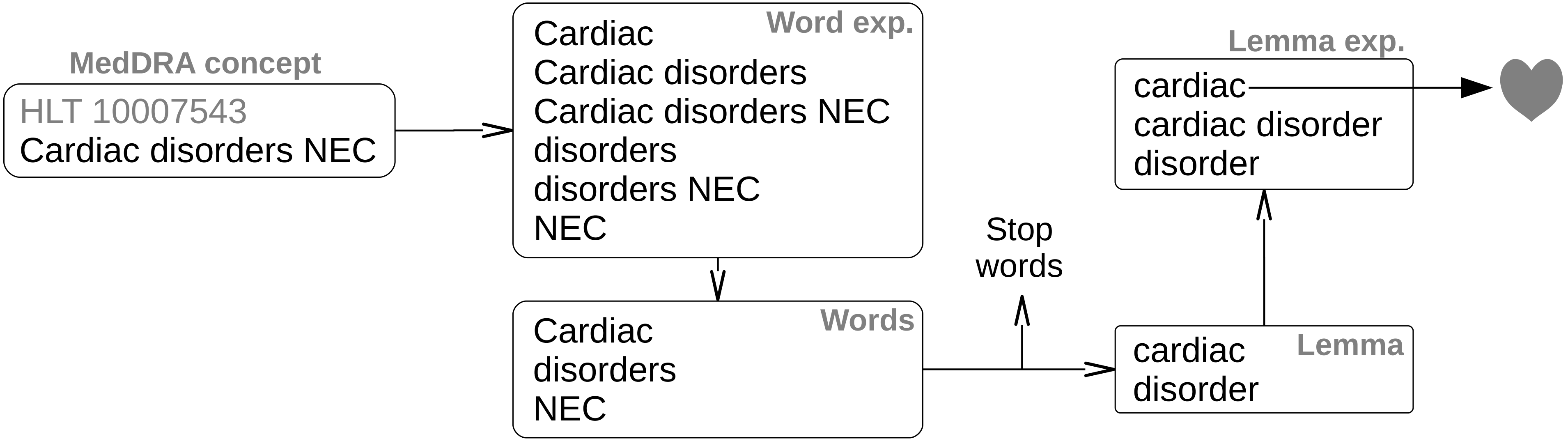
- **Combinatorial grammars**
  - 150 pictograms
  - 5 colors
  - 30 shapes
- **=> thousands of icons**
- **A formal semantics (based on an OWL 2.0 ontology)** [KBS]

Lexical methods

- **Design of an OWL ontology**
  - MedDRA terms with codes, labels, parent-child relations
  - Labels are decomposed in words and lemmas



MedDRA concept
has for word expression
Word expression
has for lemma expression
Lemma expression
mapped to
VCM concept
child of / parent of
Word
lemmatized form
Lemma
VCM pictogram
*
*
*
*
*
*
*
*
*
*
1



MedDRA concept
HLT 10007543
Cardiac disorders NEC

Word exp.
Cardiac
Cardiac disorders
Cardiac disorders NEC
disorders
disorders NEC
NEC

Words
Cardiac
disorders
NEC

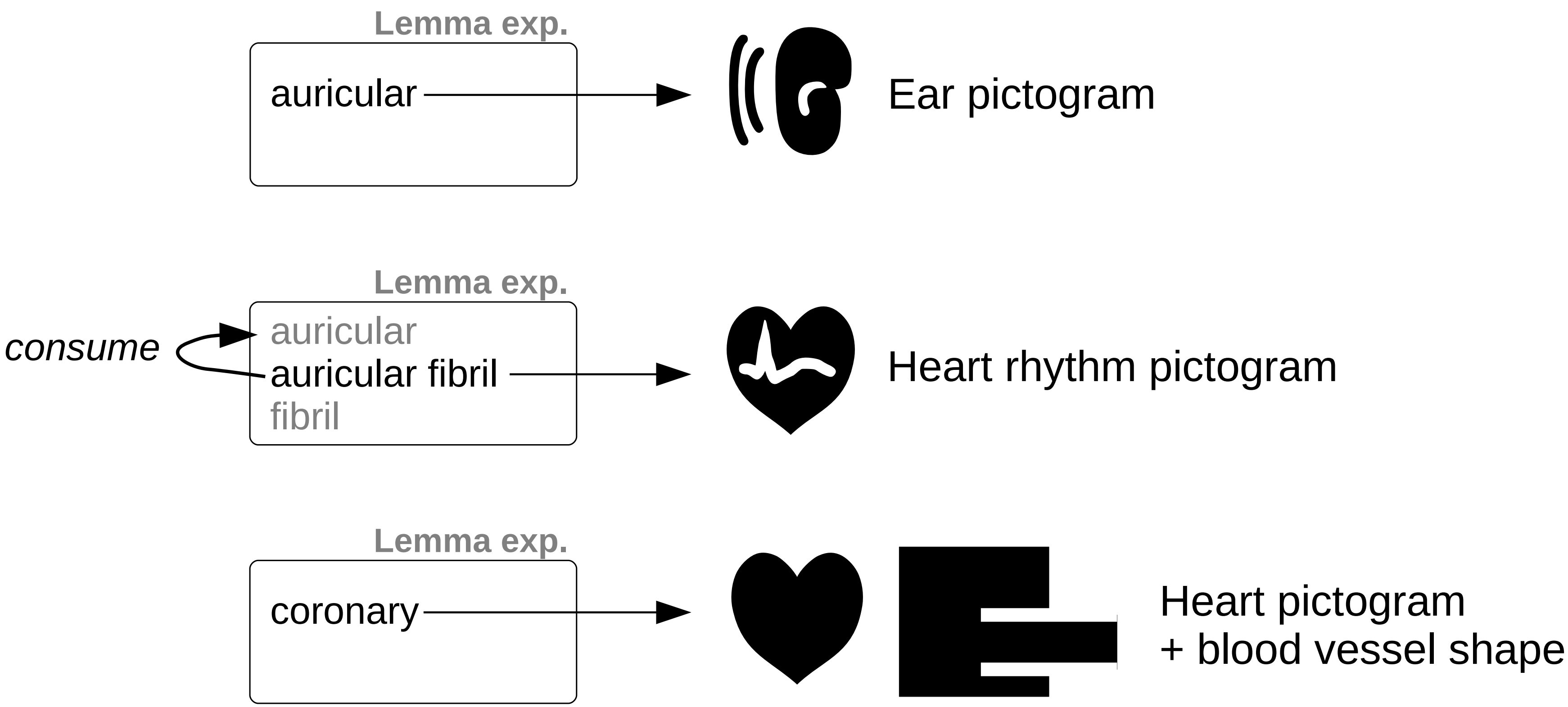
Stop words

Lemma
cardiac
disorder

Lemma exp.
cardiac
cardiac disorder
disorder

Lexical methods

- **OWL ontology**
  - Association between lemma expressions and VCM concepts



Lemma exp.

auricular

Ear pictogram

Lemma exp.

consume

auricular

auricular fibril

fibril

Heart rhythm pictogram

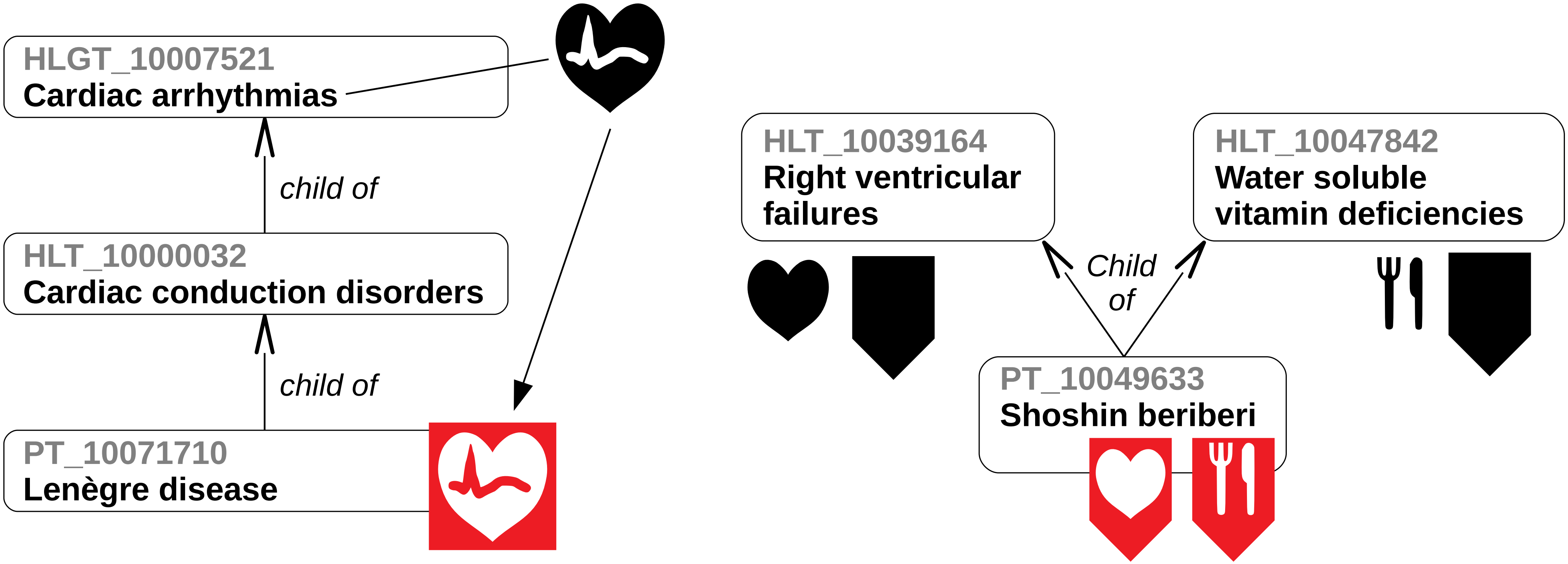
Lemma exp.

coronary

Heart pictogram
+ blood vessel shape

Semantic methods

- Based on multiple inheritance through the MedDRA multiaxial classification



HLGT_10007521
Cardiac arrhythmias

child of

HLT_10000032
Cardiac conduction disorders

child of

PT_10071710
Lenègre disease

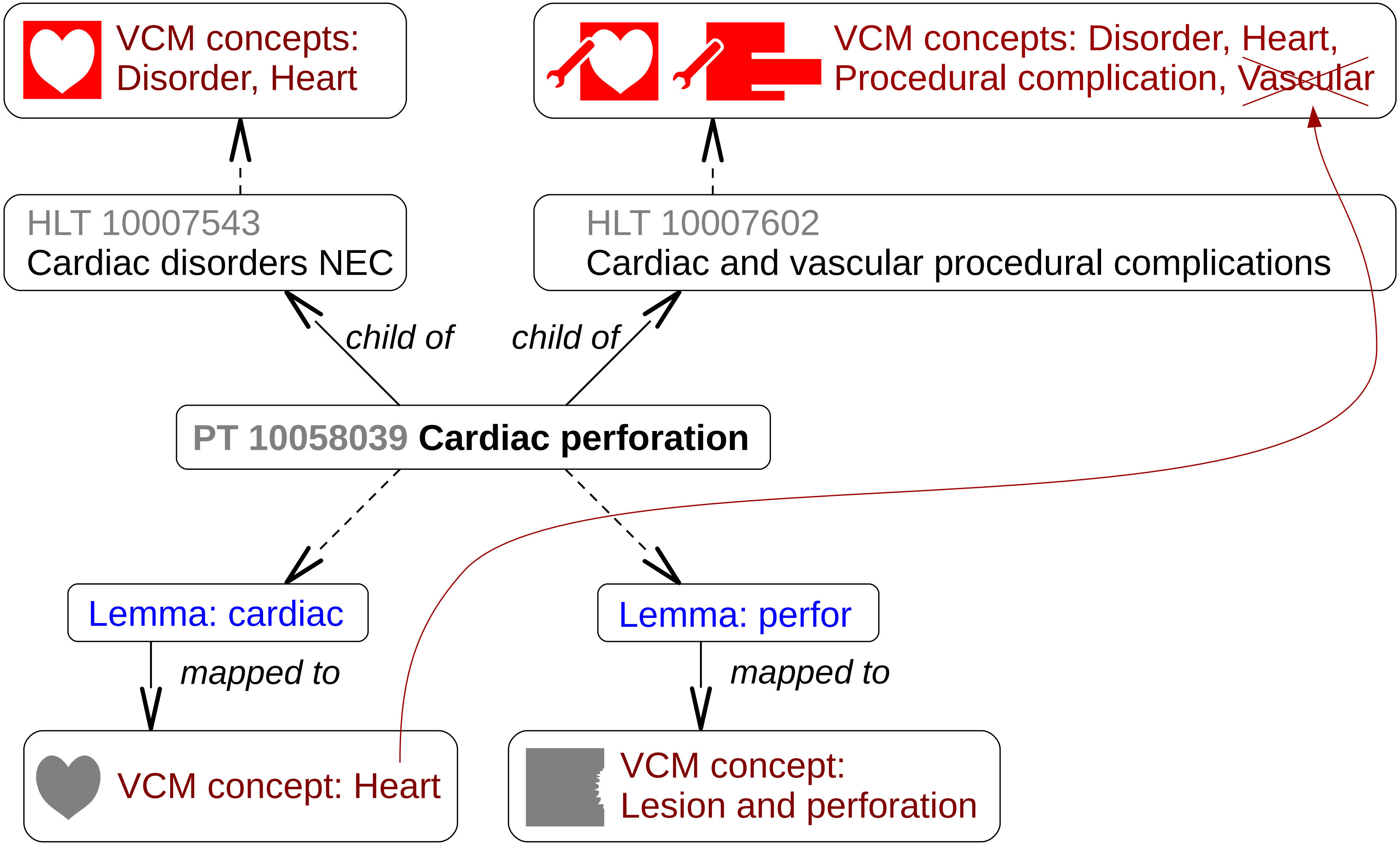
HLT_10039164
Right ventricular failures

HLT_10047842
Water soluble vitamin deficiencies

Child of

PT_10049633
Shoshin beriberi

Combining both methods



VCM concepts: Disorder, Heart

VCM concepts: Disorder, Heart, Procedural complication, ~~Vascular~~

HLT 10007543 Cardiac disorders NEC

HLT 10007602 Cardiac and vascular procedural complications

PT 10058039 Cardiac perforation

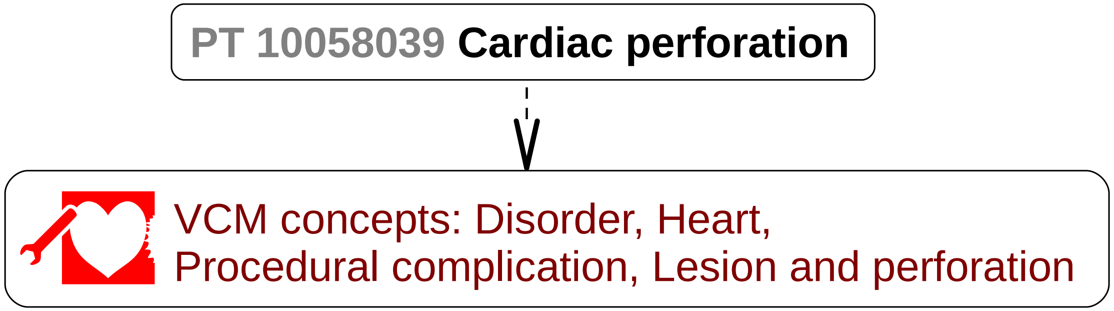
Lemma: cardiac

Lemma: perfor

VCM concept: Heart

VCM concept: Lesion and perforation

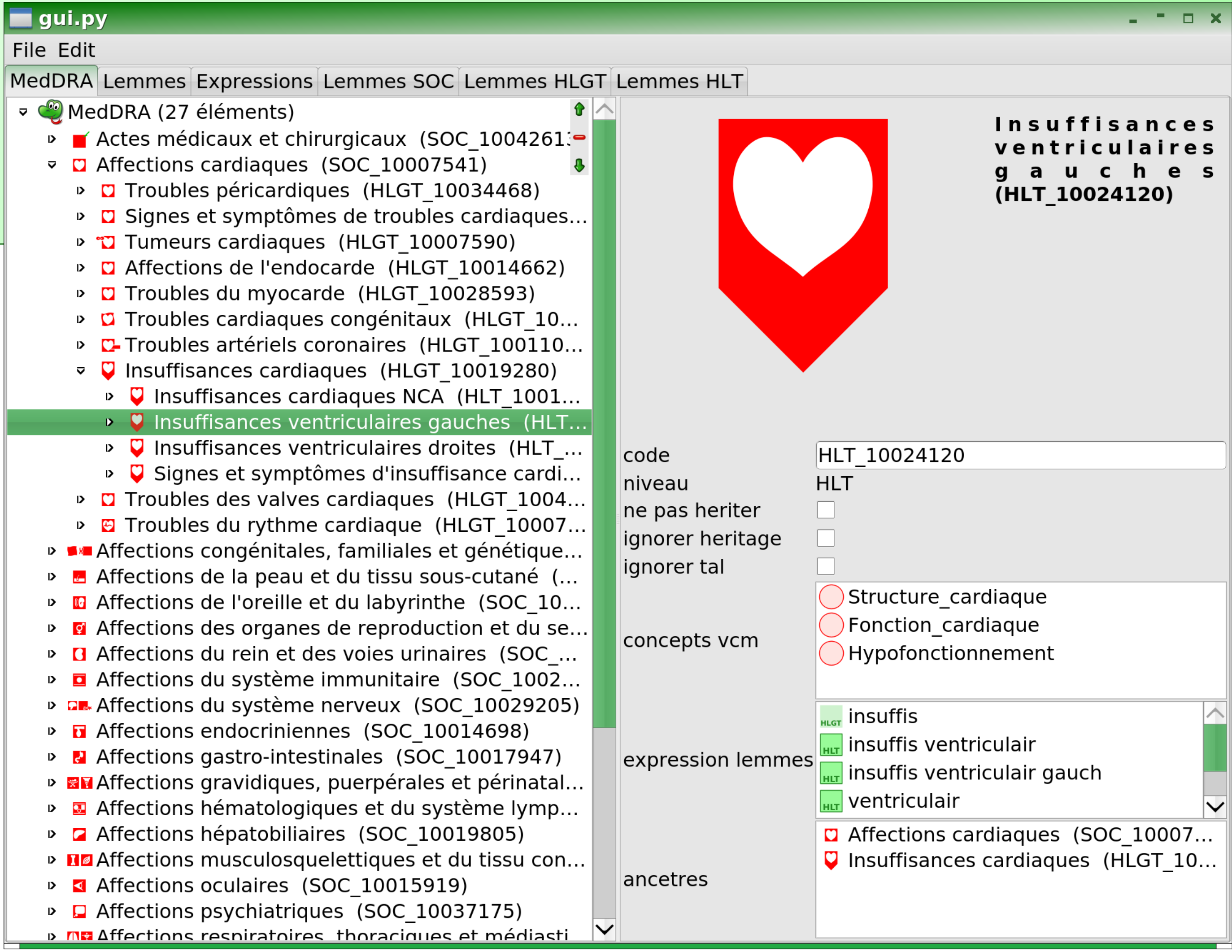
Combining both methods



PT 10058039 **Cardiac perforation**

VCM concepts: Disorder, Heart,
Procedural complication, Lesion and perforation

Results



gui.py

File Edit

MedDRA | Lemmes | Expressions | Lemmes SOC | Lemmes HLGT | Lemmes HLT

- MedDRA (27 éléments)
  - Actes médicaux et chirurgicaux (SOC_1004261…
  - Affections cardiaques (SOC_10007541)
    - Troubles péricardiques (HLGT_10034468)
    - Signes et symptômes de troubles cardiaques…
    - Tumeurs cardiaques (HLGT_10007590)
    - Affections de l'endocarde (HLGT_10014662)
    - Troubles du myocarde (HLGT_10028593)
    - Troubles cardiaques congénitaux (HLGT_10…
    - Troubles artériels coronaires (HLGT_100110…
    - Insuffisances cardiaques (HLGT_10019280)
      - Insuffisances cardiaques NCA (HLT_1001…
      - Insuffisances ventriculaires gauches (HLT…
      - Insuffisances ventriculaires droites (HLT_…
      - Signes et symptômes d'insuffisance cardi…
    - Troubles des valves cardiaques (HLGT_1004…
    - Troubles du rythme cardiaque (HLGT_10007…
  - Affections congénitales, familiales et génétique…
  - Affections de la peau et du tissu sous-cutané (…
  - Affections de l'oreille et du labyrinthe (SOC_10…
  - Affections des organes de reproduction et du se…
  - Affections du rein et des voies urinaires (SOC_…
  - Affections du système immunitaire (SOC_1002…
  - Affections du système nerveux (SOC_10029205)
  - Affections endocriniennes (SOC_10014698)
  - Affections gastro-intestinales (SOC_10017947)
  - Affections gravidiques, puerpérales et périnatal…
  - Affections hématologiques et du système lymp…
  - Affections hépatobiliaires (SOC_10019805)
  - Affections musculosquelettiques et du tissu con…
  - Affections oculaires (SOC_10015919)
  - Affections psychiatriques (SOC_10037175)
  - Affections respiratoires, thoraciques et médiasti…

Insuffisances ventriculaires gauches (HLT_10024120)

code	HLT_10024120
niveau	HLT
ne pas heriter	☐
ignorer heritage	☐
ignorer tal	☐
concepts vcm	Structure_cardiaque Fonction_cardiaque Hypofonctionnement
expression lemmes	insuffis insuffis ventricular insuffis ventricular gauch ventriculair
ancetres	Affections cardiaques (SOC_10007… Insuffisances cardiaques (HLGT_10…

- **User interface**
  - For mapping lemma expressions to VCM concepts
  - Python 3
  - Use OwlReady2 for accessing the OWL ontology [AIM]

OwlReady2

Ontology-oriented programming in Python [AIM]

```
from owlready2 import *

onto_termino    = get_ontology("http://test.org/meddra_vcm_tal.owl")
onto_alignement = get_ontology("http://test.org/meddra_vcm_alignement.owl")

with onto_termino:

  class ConceptTermino(Thing):
    def __repr__(self):
      c = MEDDRA[self.code]
      if c.meddra_type == "PT":
        return "%s  (%s %s)" % (c.term, c.code, c.primary_soc.term)
      return "%s  (%s)" % (c.term, c.code)

    def _get_concepts_vcm_her(self):
      vcms_her = set()
      for parent in self.parents:
        vcms_her.update(parent._get_concepts_vcm_a_heriter())
      return vcms_her

    def _get_concepts_vcm_tal(self):
      vcms_tal = set()
      expressions = [expression_mot.forme_lemmatise
                     for expression_mot in self.expression_mots
                     if  expression_mot.forme_lemmatise.aligne_avec_concepts_vcm or expression_mot.forme_

      expressions = expressions_prioritaires(expressions)
      for expression in expressions:
        vcms = expression.aligne_avec_concepts_vcm

        vcms_tal.update(vcms)
      return vcms_tal
```

Results

- **Application of the methods on the cardiac SOC of MedDRA**
  - 634 MedDRA terms (excluding LLT)
    - 212 lemma expressions
      - 123 with 1 lemma
      - 76 with 2 lemmas
      - 13 with more
    - **=> 212 lemma expressions mapped in lieu of 634 (longer) terms**
  - mapped to 114 different VCM icons
    - 541 to a single icon
    - 85 to 2 icons
    - 8 to 3 icons

- **Evaluation on 50 randomly-chosen terms**
  - A medical expert mapped the terms to VCM, blindly (*gold standard*)
    - For 40 terms, the expert chose exactely the same icons
    - For 6 terms, the generated icons were incomplete or more general
    - For 4 terms, the generated icons were discordant
      - E.g. mycoplasma infections classified as fungal infections

Discussion

- **Four main approaches for mapping medical terminologies :**
  - Manual mapping
    - Long, tedious, and often not reproducible
  - Chaining existent mapping:
    - MedDRA → SNOMED CT + SNOMED CT → VCM **=>** MedDRA → VCM
    - But cumules the errors and imprecisions of each mapping
  - Lexical approach
    - Difficult with icons
    - Bag of words *vs* expressions
  - Semantic approach
    - Ontology alignment methods
    - Requires terminologies having a formal semantics

Discussion

- **The proposed method is easier than a manual mapping**
  - Lemma expressions are shorter than terms, and less numerous
- **Perspectives**
  - Extending the methods with other approaches :
    - Learning method: try to learn new lexical mapping from the already asserted ones
    - Chaining method (using SNOMED CT as an intermediate terminology) : OntoADR
  - Application of the methods to the entire MedDRA terminology
  - Integration of VCM icons in pharmacovigilance software
  - Reuse of the lemma expressions - VCM concepts mapping
    - For mapping with other terminologies, *e.g.* ICD10
    - For producing icons from free text

References

[BMC] : Lamy JB, Duclos C, Bar-Hen A, Ouvrard P, Venot A. An iconic language for the graphical representation of medical concepts. BMC Medical Informatics and Decision Making 2008;8:16

[MEDINFO] : Lamy JB, Tsopra R, Venot A, Duclos C. A Semi-automatic Semantic Method for Mapping SNOMED CT Concepts to VCM Icons. Stud Health Technol Inform 2013;192:42-6

[KBS] : Lamy JB, Soualmia LF. Formalization of the semantics of iconic languages: An ontology-based method and four semantic-powered applications. Knowledge-Based System 2017;135:159-179

[AIM] : Lamy JB. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. Artif Intell Med 2017;80:11-28