

# Annotation scheme for the MERLOT French Clinical Corpus

The development of an annotated French Clinical Corpus using the scheme below is intended as resource to 1/ provide a structured analysis of clinical documents in the corpus, and 2/ provide training and test material for automatic tools performing a similar analysis in the form of entity recognition and relation extraction.

## 1. Definitions and examples for each annotation scheme category, modality, relation and event

### 1.1 Entities

The annotation scheme for entities was derived in part from the UMLS Semantic Groups, described in [1] and [2], and the annotation scheme used in [3]. Terms followed by '\*' are UMLS Semantic Groups. Terms in *Italic* are UMLS Semantic Types. Below are definitions and examples for each category in the scheme.

#### **Anatomy\***

UMLS definition Any part or component of the body. Includes the following UMLS semantic types: [Anatomical Structure](#), [Body Location or Region](#), [Body Part Organ or Organ Component](#), [Body Space or Junction](#), [Body Substance](#), [Body System](#), [Cell](#), [Cell Component](#), [Embryonic Structure](#), [Fully Formed Anatomical Structure](#), [Tissue](#)

However, terms referring to localization in the body without reference to a specific anatomic part are not annotated.

#### Examples

- (annotate) *anus*; *pied*; *Placenta*, *Artère fémorale droite*
- (do NOT annotate) *périphérique*; *à droite*

#### **Biological Process or Function**

Definition A process or state which occurs naturally or as a result of an activity. Biological Process or Function includes the following UMLS semantic types: [Biologic Function](#); [Cell Function](#); [Genetic Function](#); [Molecular Function](#); [Natural Phenomenon or Process](#); [Organ or Tissue Function](#); [Organism Function](#); [Physiologic Function](#)

Examples *Transit*;

## Chemicals and drugs

(Webster) Definition matter of particular or definite chemical constitution; a substance used as a medication or in the preparation of medication. Includes the following UMLS semantic types: [Antibiotic](#); [Biomedical or Dental Material](#); [Carbohydrates](#); [Chemical](#); [Chemical Viewed Functionally](#); [Chemical Viewed Structurally](#); [Clinical Drug](#); [Hazardous or Poisonous Substance](#); [Inorganic Chemical](#); [Pharmacological Substance](#); [Vitamin](#).

Caveat Please note that some *Biologically Active Substances* including *Enzymes*, *Immunologic Factors* and *Receptors* may be better categorized as *Genes and Proteins*. Any substance listed in Entrez Gene may be categorized as *Genes and Proteins*

Examples [Questran](#); [Insuline](#); [corticoïdes](#); [formol](#); [traitement anti-viral](#); [traitement médicamenteux](#)

## Concept and Ideas\*

Definition an abstract or generic idea generalized from particular instances. Includes the following UMLS semantic types: [Classification](#), [Conceptual Entity](#), [Functional Concept](#), [Group Attribute](#), [Idea or Concept](#), [Intellectual Product](#), [Language](#), [Qualitative Concept](#), [Quantitative Concept](#), [Regulation or Law](#), [Spatial Concept](#). Note that entities of type [Temporal Concept](#) are annotated with entity type *Temporal*.

Examples [terme](#); [Longueur](#); [Poids](#)

## Devices \*

UMLS definition A manufactured object used primarily in the diagnosis, treatment, or prevention of physiologic or anatomic disorders; A manufactured object used primarily in carrying out scientific research or experimentation. Devices include the following UMLS Semantic Types: [Drug Delivery Device](#); [Medical Device](#); [Research Device](#)

Examples [Pompe à insuline](#), [sonde](#), [prestige VH](#) [General Electric](#), [tube](#), [flacon](#), [spray](#)

## Disorders

!! [Signs or Symptoms](#) are listed separately !!

Webster Definition (disease): a condition of the living animal or plant body or of one of its parts that impairs normal functioning and is typically manifested by distinguishing signs and symptoms. Disorders include the following UMLS semantic types: [Acquired Abnormality](#); [Anatomical Abnormality](#); [Cell or Molecular Dysfunction](#); [Congenital Abnormality](#); [Disease or Syndrome](#); [Experimental Model of Disease](#); [Injury or Poisoning](#); [Mental or Behavioral Dysfunction](#); [Pathologic Function](#); [Neoplastic Process](#); some [Findings](#) may also be relevant.

Examples [Diabète](#), [insuffisance mitrale](#), [MFIU](#), [fracture](#), [éventration](#)

Annotate the more specific entity and including the longest span as possible, even though the full

entity is not found in any terminological resource: e.g. [décompensation d'un diabète de type 2](#)

## Genes and Proteins

Definition A gene is defined as the section of DNA that represents the blueprint for the construction of a protein. Usually, a gene and the protein it encodes are referred to by similar names. This category also includes the following UMLS semantic types: [Amino Acid, Peptide or Protein; Enzyme, Lipid; Immunologic Factor; Indicator, Reagent, or Diagnostic Aid; Gene or Genome; Nucleic Acid, Nucleoside or Nucleotide; Receptor](#). To summarize, any substance listed in Entrez Gene may be categorized as *Genes and Proteins*

Examples [PTX1; POLYSERASE 3; DUARTE BRAIN-SPECIFIC PROTEIN](#)

## Hospital

Definition The name of a health care facility, office or ward.

Examples [Unité Henri Mondor; Hopital de la Croix Rouge; CHU de Rouen](#)

## Living Beings

Definition An individual form of life that is not human (see separate entity *Person*), such as a plant, animal, bacterium, protist, or fungus; a body made up of organs, organelles, or other parts that work together to carry on the various processes of life. Includes the following UMLS semantic types: [Alga; Amphibian; Animal; Archeon; Bacterium; Bird; Fish; Fungus; Invertebrate; Mammal; Organism; Plant; Reptile; Rickettsia or Chlamydia; Vertebrate; Virus](#)

Examples [chien, salmonelle](#)

## Medical procedures

Definition An activity relating to the practice of medicine or involving the care of patients, including diagnosis or treatment procedures, techniques or methods. It includes the following UMLS semantic types: [Diagnostic Procedures; Health Care Activity; Laboratory Procedure; Therapeutic or Preventive Procedure](#).

Examples [Consultation psychiatrique; Angiographie; Cholécystectomie, IMG, traitement anti-viral, traitement médicamenteux;](#)

Generic medical procedures should also be annotated.

Examples [Traitement, prélèvement](#)

Some section heads should be annotated as MedicalProcedure:

Examples [Antécédents; anamnèse → C0025084 Medical History Taking ; Examen clinique → C0031809 Physical Exam](#)

## Persons

Definition For Human Living Beings. Corresponds to the Semantic Type [Human](#).

Examples [patiente](#), [Docteur Dupond](#)

Include honorifics : [Monsieur le Docteur Durand](#)

Annotate endearment/formal references to persons as one entity : [Mon cher confrère](#)

Discontinuous entity in form-like presentation of patient : NOM: [Durand](#) PRENOM: [Michel](#)

## Sign or Symptom

Definition An observable manifestation of a disease or condition based on clinical judgment, or a manifestation of a disease or condition which is experienced by the patient and reported as a subjective observation. Includes the UMLS Semantic Type [Sign or Symptom](#); some [Findings](#) may also be relevant.

Examples [Fatigue](#); [Douleurs](#); [ballonnement](#); [anneau fibreux](#); [prise d'activité](#); [prise de poids](#)

[Pas de signe d'insuffisance cardiaque](#) : signe = SignOrSymptom (SOSY), insuffisance cardiaque = DISO

## Attribute entities

### Aspect

Definition An entity that represents a change (movements of object are not covered).

Examples

- (annotate) [relai](#), [rechute](#), [augmenter](#), [arrêter](#), [continuer](#), [progression de la maladie](#)
- (do not annotate) [la progression de l'endoscope est facile](#)

### Assertion

Definition a phrase or text span that provides motivation for assigning a given Modality to an Entity.

Examples

- [pas de](#) (indicates negation)
- [suspecté](#), [éventuel](#), [semble](#) (indicates possibility)
- [si](#) (indicates SubjectToCondition)
- [présence de](#) (indicates presence)
- [Recherche](#) de DISORDER ( indicates possibility)

## Drug Attributes

- **AdministrationRoute** : administration route of the medication
  - Examples injection, oralement, per os, IV
- **Dosage** : dosage of the prescription
  - Examples deux gouttes, 1 cachet, 10 mg (as in 10 mg/jour), 1.1.1 (as in MOTILIUM : 1.1.1)
- **DrugForm** : form of the medication
  - Examples sirop, pilule, sachet, comprimé, crème
- **Strength** : strength of the drug (such as Doliprane 500)
  - Examples 10 mg/ml, 20 (as in INEXIUM 20), 200 mg (as in OFLOCET 200 mg x3/jour)

## Localization

Definition precise area where an entity is located (e.g. body side).

Examples à droite, bilatéral

## Measure

Webster Definition a figure, extent, or amount obtained by measuring or observing. Measurement entities are typically composed of a value and/or unit entity. They also include subjective qualifications of the shape, color, or other attributes of measured entities.

Typically, for quantitative measures, the annotation should include a value and the corresponding unit. For qualitative measures, the annotation often consists of an adjective or adverb.

Examples 37.2°C, 20 cm, quelques [ Quantitative measure ] normale, opaque, oblong, sévère, lentement [ Qualitative measure ]

## Temporal

Definition temporal expressions, such as times, dates, durations, etc. Includes UMLS Semantic Type *Temporal Concept*. Has 4 modalities used to specify the type (see Modalities section).

Examples 23 novembre 1996, pendant deux semaines, deux fois par jour, aujourd'hui, il y a trois semaines, 15 SA, vespéral, période post-prandiale, récent[DATE], récemment[DATE], rare[FREQ]

## 1.2 Modalities or types

For all Entities except Assertion:

### Ambiguous

Definition an entity is considered *ambiguous* if there is a question about whether an annotation should be made at all.

Examples In the phrase *Vitals: T: 100.5F* there can be a question of whether to annotate *T: 100.5F* as a Disorder (fever), or not annotate at all.

(Note that the correct annotation in this case is **T** as **CONC** and **100.5F** as **MEAS[quantitative]**)

## DocTime

Definition Time relation related to the current medical visit. DocTime attribute applies primarily to Event entities, i.e. [Disorder](#), [SignOrSymptom](#), [MedicalProcedure](#), [Chemical\\_Drugs](#), [BiologicalProcessOrFunction](#), [Concept\\_Idea](#)

Values: Before, Overlap (default), After, Before\_overlap

The default value is assumed to be *Overlap* and does not need to be annotated.

### Examples

- Recherche de **DISORDER** → Overlap. The disorder is suspected: if confirmed the patient experiences it at the time of suspicion.
- Recherche[**PROC**] de **DISORDER** → After + Recherche [ASRT]. Exams conducted to evidence the disorder will occur in the future.
- Indication: patient **diabétique** → Before\_Overlap. In general, indication means that the patient is seen for a problem that has been previously diagnosed, and is still currently experienced by the patient.
- Antécédents → Before or → Before overlap depending on the type of disease/procedure; DNID/HTA is likely still experienced by the patient (Before\_overlap), AVC/appendectomie is likely a one-time event → Before
- Events evidenced by a current exam → Before\_overlap: Existence d'une **aérobilie** → Before\_Overlap; Conclusion: **maladie hémorroïdaire de stade 3** → Before\_Overlap
- Je revois Mme Martin pour le suivi de son **cancer du colon** → Before\_Overlap.
- Mr. Martin, **né** le 1 janvier 1980 → Before
- La dernière **endoscopie** montre... → Before
- Prochaine **hospitalisation** programmée/prévue le...
  - (do not annotate) **prochaine** + After + programmée/prévue [ASRT] (présence)
- Une **hospitalisation** est proposée... → After + proposé[ASRT]

## Abbreviations

Webster definition a shortened form of a written word or phrase used in place of the whole.

An entity is considered an abbreviation is the entire span of the entity corresponds to an abbreviated form

NB : the default value is *no*, so you only need to use this modality to indicate that the entity IS an abbreviated form (yes value).

### Examples

- (annotate) **EEG**; **DNID**; **IMG**;
- (do not annotate) **7 kg**; **cirrhose** **OH**

## Coreference (CorefPronoun)

Definition indicates that an entity is a pronominal coreference

NB : the default value is *no*, so you only need to use this modality to specify that there is a coreference (yes value).

Do not use for lexical coreference (e.g. *patiente, traitement*)

Examples *il; elle; ceci; celui-ci; l; son; sa*

## Modalities for Measure

Specify the type of measure:

- **Quantitative** : a measure is considered quantitative if it is expressible in terms of a value and/or unit
  - Examples *20 cm, 37 degrés, PH 4*
- **Qualitative** : a measure is considered qualitative if it relies on a judgment or appreciation
  - Examples *normale, opaque, abimé*

## Modalities for Persons

Specify the type of person:

- Patient
- Family
- Donor
- Healthcare professional
- Other

## Modalities for Temporal expressions

Specify the type of temporal expression:

- **Date** : describes a calendar date (can be relative)
  - Examples
    - (annotate) *1981, 25/12/99, aujourd'hui, il y a trois semaines, récemment*
    - (do not annotate) *date de sortie, date de reprise d'activité*
- **Time** : specific time points within a day
  - Examples *16h55, trois heures de l'après-midi*
- **Duration** : reflects a span of time
  - Examples *depuis six mois, une heure, cinq ans, 15 SA*
- **Frequency** : describes a set of times
  - Examples *deux fois par jour, toutes les trois heures, /jour (as in 10 mg/jour), rare*

All other modalities will be treated as a relation between the entity the modality is relevant for and the textspan used a motivation for assigning the modality.

## 1.3 Events

A subset of entities are considered as events and will participate in temporal relations (i.e. with another event or a temporal expression) and have DocTime attributes. These entities are:

- Disorder
- SignOrSymptom
- MedicalProcedure
- Chemicals\_Drugs
- Concept\_Idea
- BiologicalProcessOrFunction

## 1.4 Relations

- Do not annotate all relations except those **between the entities that appear syntactically closer in the sentence**
- Location\_of: annotate the more precise entities
- Mark all relations Location\_of even if other relations co-occur between entities

The annotation scheme for relations was derived in part from the UMLS Semantic Network [http://www.nlm.nih.gov/research/umls/META3\\_current\\_relations.html](http://www.nlm.nih.gov/research/umls/META3_current_relations.html), from the SHARP template annotation guidelines [4], and from the fourth i2b2/VA Shared-Task and Workshop <https://www.i2b2.org/NLP/Relations>.

UMLS and i2b2-based relations (UMLS relations followed by \*)

### Affects

Disorder|SignOrSymptom|Chemicals\_Drugs|MedicalProcedure Affects BiologicalProcessOrFunction

Definition Produces a direct effect on. Implied here is the altering or influencing of an existing condition, state, situation, or entity. This includes has a role in, alters, influences, predisposes, catalyzes, stimulates, regulates, depresses, impedes, enhances, contributes to, leads to, and modifies.

### Causes \*

Living Beings Causes Disorder

Chemicals\_Drugs Causes Disorder

MedicalProcedure Causes Disorder

Disorder Causes Disorder

SignOrSymptom Causes Disorder

Definition Brings about a condition or an effect. Implied here is that an agent, such as for example, a



pharmacologic substance or an organism, has brought about the effect. This includes induces, effects, evokes, and etiology.

## Complicates \*

Disorder Complicates Disorder

Chemicals\_Drugs Complicates Disorder

MedicalProcedure Complicates Disorder

Definition Causes to become more severe or complex or results in adverse effects.

### Examples

- Hépatite C [DISO] d'évolution cirrhogène [DISO] → cirrhogène *Complicates* Hépatite C
- sténose [DISO] inflammatoire [DISO] → inflammatoire *Complicates* sténose

## Conducted

MedicalProcedure Conducted for Disorder

Definition When a test is conducted to investigate a Disorder and the outcome is unknown/does not result in a diagnosis.

Example ASP réalisé au vue d'une distention gastrique.

## Experiences

Person Experiences Disorder

Person Experiences SignOrSymptom

Person Experiences MedicalProcedure

Person Experiences Chemical\_Drugs

Person Experiences BiologicalProcess

Person Experiences Concept\_Idea

Definition When a Person (e.g. patient) is affected by a Disorder, SignOrSymptom; when a Person (e.g. patient) is subjected to a MedicalProcedure. Events planned in the future, e.g. medical exams should be annotated.

### Examples

[Patient] présentant un [éthylisme chronique] : Relation Experiences between Persons *Patient* and Disorder *éthylisme chronique*

[II] a eu une [résection endoscopique] : Relation Experiences between Persons // and MedicalProcedure *résection endoscopique*

Le [patient] est sous [antibiotiques] : Relation Experiences between Persons *Patient* and Chemical\_drugs *antibiotiques*

### **Interacts\_with \***

Chemicals\_Drugs Interacts\_with Chemicals\_Drugs

Definition Acts, functions, or operates together with.

### **Localization\_of**

Localization Localization\_of Anatomy

Localization Localization\_of Disorder

Localization Localization\_of SignOrSymptom

Localization Localization\_of MedicalProcedure

Definition The spatial or relative localization of an entity.

Examples :

[adénocarcinome rectal [bifocal]] : relation Localization\_of between Localization *bifocal* and Disorder *adénocarcinome rectal bifocal*

### **Location\_of**

Anatomy Location\_of Anatomy

Anatomy Location\_of Disorder

Anatomy Location\_of SignOrSymptom

Anatomy Location\_of MedicalProcedure

Anatomy Location\_of LivingBeings

Anatomy Location\_of Persons

Hospital Location\_of Persons

Hospital Location\_of MedicalProcedure

Definition The position, site, or region of an entity or the site of a process.

## Examples

[*fracture du [rocher droit]*] : relation Location\_of between Anatomy *rocher droit* and Disorder *fracture du rocher droit*

[*échographie [abdominale]*] : relation Location\_of between Anatomy *abdominale* and MedicalProcedure *échographie abdominale*

[*examen*] pratiqué au [CHU] : relation Location\_of between MedicalProcedure *examen* and Hospital *CHU*

le [*patient*] a été hospitalisé au [CHU] : relation Location\_of between Person *patient* and Hospital *CHU*

## Measure\_of

Measure Measure\_of {Entity}

Definition The quantitative or qualitative result of a medical procedure such as lab test or physical examination.

## Examples

[*délirium tremens*] [*sévère*] : Relation Measure\_of between Measure *sévère* and Disorder *délirium tremens*

[*Température*] : [*37°4*] : Relation Measure\_of between Measure *37°4* and Concept\_Idea *Température*

## Performs

Person Performs MedicalProcedure

## Examples :

[*Consultation psychiatrique*] avec le [*Dr. House*] : Relation Performs between MedicalProcedure *consultation psychiatrique* and Person *Dr. House*

## Physically Related To \*

CONC Physically\_related\_to Anatomy|Disorder|Person

Anatomy Physically\_related\_to Anatomy

Definition Related by virtue of some physical attribute or characteristic.

## Examples

- *calibre*[CONC] du *Cholédoque*[ANAT] → *calibre* Physically\_related\_to *cholédoque*
- *taille*[CONC] de la *tumeur*[DISO] → *taille* Physically\_related\_to *tumeur*

**Prevents \***

Chemicals\_Drugs Prevents Disorder|SignOrSymptom

Device Prevents Disorder|SignOrSymptom

MedicalProcedure Prevents Disorder|SignOrSymptom

Definition Stops, hinders or eliminates an action or condition.

**Reveals**

MedicalProcedure|SignOrSymptom Reveals Disorder

MedicalProcedure Reveals SignOrSymptom

Definition When a test is conducted and the outcome is known/leads to a diagnosis.

Examples

- (annotate) [RX] septembre 2006 : [Chondrolyse] : Relation Reveals between MedicalProcedure RX and Disorder *Chondrolyse*
- (annotate) A l'ASP, pas d'opacité de tonalité calcique → Reveals between MedicalProcedure ASP and (negated) SignOrSymptom *opacité de tonalité calcique*
- (do not annotate) ASP pour distention gastrique → see Conducted

**Treats \***

Chemicals\_Drugs Treats Disorder|SignOrSymptom

MedicalProcedure Treats Disorder|SignOrSymptom

Devices Treats Disorder|SignOrSymptom

Definition Applies a remedy with the object of effecting a cure or managing a condition.

Examples

une [PR] très agressive ayant nécessité plusieurs [interventions] : Relation Treats between MedicalProcedure *interventions* and Disorder *PR*

[Hypertension artérielle] ayant motivé l'instauration d'un traitement par [TENORMINE] : Relation Treats between Chemical\_Drugs *TENORMINE* and Disorder *Hypertension artérielle*

[suivi] des [douleurs gastriques] : Relation Treats between Procedure *suivi* and Disorder *douleurs gastriques*

## Used\_for

Devices Used\_for Chemicals\_Drugs

Devices Used\_for Procedures

Devices Used\_for LivingBeings|Person

Chemicals\_Drugs Used\_for Procedures

Examples :

L' [examen] a été effectué avec un [appareil] décontaminé : Relation Used\_for between Device *appareil* and MedicalProcedure *examen*

[VENTOLINE] [SPRAY] : Relation Used\_for between Device *SPRAY* and Chemicals\_Drugs *VENTOLINE*

[Traitement] par [corticoïdes] : Relation Used\_for between Chemicals\_Drugs *corticoïdes* and MedicalProcedure *traitement*

## Drug-attribute relations

### HasAdministrationRoute

Chemicals\_Drugs HasAdministrationRoute AdministrationRoute

Examples

[GENTAMYCINE] 250 mg en [IV] : Relation HasAdministrationRoute between Chemicals\_Drugs *GENTAMYCINE* and Route *IV*

relai [per os] par [OROKEN] : Relation HasAdministrationRoute between Chemicals\_Drugs *OROKEN* and Route *per os*

### HasDosage

Chemicals\_Drugs HasDosage Dosage

Examples

[TENORMINE] : [1 cp]/jour : Relation HasDosage between Chemicals\_Drugs *TENORMINE* and Dosage *1 cp*

[FOLDINE] : [1/1/1] : Relation HasDosage between Chemicals\_Drugs *FOLDINE* and Dosage *1/1/1*

### HasDrugForm

Chemicals\_Drugs HasDrugForm DrugForm

## Examples

[TENORMINE] : 1 [cp]/jour : Relation HasDrugForm between Chemicals\_Drugs *TENORMINE* and DrugForm *cp*

[FORLAX] : 2 [sachets]/jour : Relation HasDrugForm between Chemicals\_Drugs *FORLAX* and DrugForm *sachets*

## HasStrength

Chemicals\_Drugs HasStrength Strength

### Examples

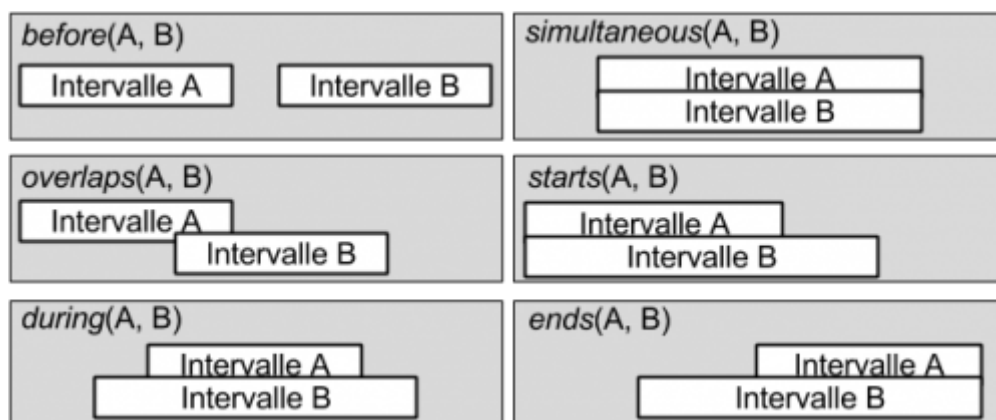
[NICOBION] [500] : Relation HasStrength between Chemicals\_Drugs *NICOBION* and Strength *500*

[OROKEN] [200 mg] x3/jour : Relation HasStrength between Chemicals\_Drugs *OROKEN* and Strength *200 mg*

## Temporal relations

Temporal relation between an event and a temporal entity, or between two events.

- **Before** : an event precedes, occurs before another event/temporal expression
- **Simultaneous** : an event happens at exactly the same time as another event/temporal expression
- **During** : the temporal span of an event is completely contained within the span of another event or temporal expression
- **Begins\_on** : the event begins on the event or temporal expression it's related to
- **Ends\_on** : the event ends on the event or temporal expression it's related to
- **Overlap** : an event happens almost at the same time, but not exactly, as another event/temporal expression



## Aspectual relations

Between an Aspect entity and an entity

- **Start** : start or initiation of an event
  - [relai] par [OROKEN] relation Start between Aspect *reprise* and Chemical\_Drugs *OROKEN*
- **Recurrence\_StartAgain** : indicates that an event starts again
  - une [récidive] de [distension urétrale] : relation Recurrence\_StartAgain between Aspect *récidive* and Disorder *distension urétrale*
  - [reprise] des [antibiotiques] : relation Recurrence\_StartAgain between Aspect *reprise* and Chemical\_Drugs *antibiotiques*
- **Increase** : indicates an increase (e.g. of a drug)
  - [augmentation] de la [corticothérapie] : relation Increase between Aspect *augmentation* and Chemical\_Drugs *corticothérapie*
- **Decrease** : indicates a decrease (e.g. of a drug)
- **Improve** : indicates an improvement (e.g. in patient condition)
  - [amélioration] du [diabète] sous insuline : relation Improve between Aspect *amélioration* and Disorder *diabète*
- **Worsens** : indicates a negative change (e.g. in patient health status)
- **Continue** : shows the continuation of an event
- **Stop** : indicates the ending of an event
  - [arrêt] des [antibiotiques] : relation Stop between Aspect *arrêt* and Chemical\_Drugs *antibiotiques*

## Assertion relations

These modalities are annotated as a relation between an Entity and an Assertion.

For all entities

Status

- **Presence**
  - [présence de] [lésions] : relation Presence between Assertion *présence de* and Disorder *lésions*
- **SubjectToCondition**
  - [si] apparition de lésions, revenir en [consultation] : relation SubjectToCondition between Assertion *si* and MedicalProcedure *consultation*
- **Possible**
  - [doute] sur [VCT] : relation Possible between Assertion *doute* and Disorder *VCT*
  - [si] apparition de [lésions], revenir en consultation : relation Possible between Assertion *si* and Disorder *lésion*
- **Negation**
  - [pas d'] [insuffisance cardiaque] : relation Negation between Assertion *pas d'* and Disorder *insuffisance cardiaque*



NB: Studies on factivity of events (i.e. the level of information expressing the factual nature of events in a text) (cf. FactBank) distinguish several axes:

- Epistemic modality: degree of certainty of the source with regard to the fact that the event is (or will be) a fact of the world  
Values: certain > probable > possible > uncertain or unspecified
- Polarity: any event is presented as positive (i.e. happening) or negative (i.e. not happening)  
Values: positive, negative or unspecified

Regarding MERLOT:

- Events are marked as entities (e.g. in the case of a medication, the medication intake is considered)
- Certain is marked with relation Presence
- Probable and possible are marked with relation Possible
- Polarity is marked with relation Negation, which may also co-occur with the previews relations (Presence or Possible).

By default, the assertion value is Presence and combines with DocTime temporal values. If a DocTime value is After, we suppose that the assertion is Presence (even though we are unsure that the event might really occur), except if any textual marker specifies the contrary.

## Coreference

Only annotate pronouns if they refer to a Person entity, using the same Person entity type as the entity they refer to and assign the modality CorefPronoun.

NB : Do not assign the CorefPronoun modality to lexical coreferences (e.g. maladie, patiente). Coreference links will be annotated later on.

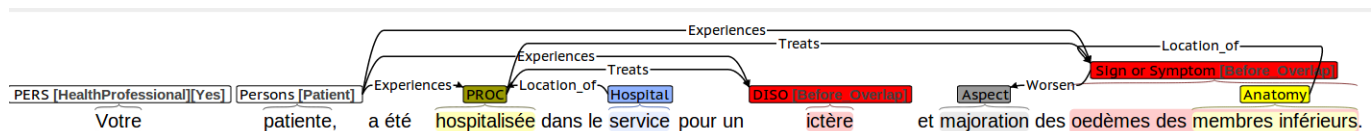
## 2. Annotation Guidelines

1. **Check before you annotate:** if you are not sure which category should be assigned to a term you are annotating, make the appropriate verification using the [Portail Terminologique de Santé](#) and the [UMLS Knowledge Source Server](#).
2. **Err on the side of caution:** if an annotation seems questionable, discard to ensure high quality annotations; or select the *Ambiguity* modality
3. **Be specific:** always annotate as specifically as possible. For instance, in the phrase *cancer du sein* the entire phrase should be annotated as a Disorder vs. the most generic *cancer*. In addition, *sein* should be annotated as an Anatomy entity. Similarly, a MedicalProcedure should encompass an anatomy entity if relevant. For instance, *examen macroscopique placentaire* should be annotated as a MedicalProcedure. In addition, *placentaire* should be annotated as an Anatomy entity.
4. **Consider context:** in some cases, a concept might belong to two categories. For example, *salmonella* may be used to refer to the bacteria (category Living Beings) or to the infection caused by the bacteria (category Disorders). Try to pick the category that seems most likely intended by the user based on the query (eg. In the phrase *salmonella treatment* the disorder seems a better fit than the bacteria).
5. **Multiple annotations:** in some cases, multiple annotations may be inferred from a single string.
  1. eg. in the phrase *un cancer du sein dépisté par une mammographie*, *cancer du sein* should be annotated as a disorder and *sein* should also be annotated as anatomy with a *location\_of* relation between *sein* and *cancer du sein*. In addition, *mammography* should be annotated as a procedure. Similarly, in the phrase *examen macroscopique placentaire* the entire phrase should be annotated as a MedicalProcedure. In addition, *placentaire* should be annotated as an Anatomy entity with a *location\_of* relation between *placentaire* and *examen macroscopique placentaire*.



2. e.g. in the phrase *carence en vitamines B6 et B12* some portions of the phrase are in fact distributed between two concepts. In this case, both *carence en vitamines B6* and *carence en vitamines ... B12* should be annotated, thus creating a discontinuous annotation for *carence en vitamines B12*.
3. e.g. in the phrase *Salmonella* it can be ambiguous whether to annotate as a Disorder (infection caused by salmonella) or as a Living Being (bacteria). If the context does not allow to make a clear decision, both annotations should be made.
6. **List of entities:** when a list of entities is encountered, each element of the list needs to be annotated separately and consistently with the other list items. For example, in the phrase *hématome déjà en voie d'organisation, clair et foncé* there is a list of qualification of the disorder entity *hématome*: the three list items *déjà en voie d'organisation, clair et foncé* have to be annotated as *Measure* entities. Note that one list item can be complex - for example *cubique ou aplati* is an item in the two-item list *cubique ou aplati, jointif*.
7. **Discontinuous annotations:** in some cases, the annotations of an entity may be inferred from several text fragments. That is the case for *carence en vitamines B12* in the phrase *carence en vitamines B6 et B12*. Discontinuous annotations may also occur in verbal phrases such as *cordons presque arrachés* where *cordons ... arrachés* should be annotated as a disorder, or *l'artère interventriculaire inférieure est très calcifiée*, where *artère ... calcifiée* should be annotated as a disorder.
8. **Misspellings:** in some cases, misspellings will appear in the texts. When the concept intended by the writer can be reasonably inferred, the annotation should be made (eg. select *cirrhose* as a disorder, even though *cirrhose* is the correct spelling).

## Sample annotations according to above scheme and guidelines



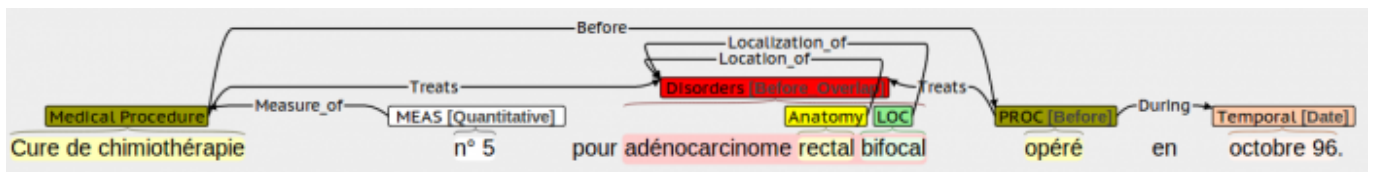
**Text** - Votre patiente a été hospitalisée dans le service pour un ictère et majoration des oedèmes des membres inférieurs.

**Translation** - 'Your patient has been admitted in the ward for icterus and worsening of lower limbs oedema'.

### Annotations

- *Votre* as Persons, with type HealthProfessional and mark of Coreferent pronoun (Yes)
- *patiente* as Persons of type Patient
- *hospitalisée* as MedicalProcedure
- *service* as Hospital
- *ictère* as Disorder with DocTime Before\_Overlap
- *majoration* as Aspect
- *oedèmes des membres inférieurs* as SignOrSymptom with DocTime Before\_Overlap
- *membres inférieurs* as Anatomy
- Experiences relation between *patiente* and *hospitalisée*
- Experiences relation between *patiente* and *ictère*
- Experiences relation between *patiente* and *oedèmes des membres inférieurs*
- Location\_of relation between *service* and *hospitalisée*
- Treats relation between *hospitalisée* and *ictère*

- Treats relation between *hospitalisée* and *oedèmes des membres inférieurs*
- Worsen relation between *oedèmes des membres inférieurs* and *majoration*
- Location\_of relation between *membres inférieurs* and *oedèmes des membres inférieurs*

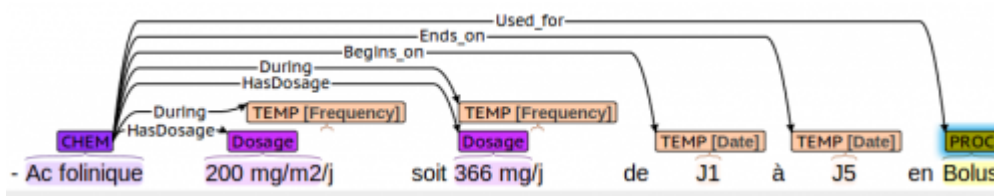


**Text** Cure de chimiothérapie n° 5 pour adénocarcinome rectal bifocal opéré en octobre 96.

**Translation** 'Chemotherapy round n° 5 for bifocal rectal adenocarcinoma. Surgical resection was performed in October 96'.

**Annotations**

- *Cure de chimiothérapie* as MedicalProcedure
- *n° 5* as Measurement (Quantitative type)
- *adénocarcinome rectal bifocal* as Disorder with DocTime Before\_Overlap
- *rectal* as Anatomy
- *bifocal* as Localization
- *opéré* as MedicalProcedure with DocTime Before
- *octobre 96* as Temporal entity of type Date
- Measure\_of relation between *n° 5* and *Cure de chimiothérapie*
- Treats relation between *Cure de chimiothérapie* and *adénocarcinome rectal bifocal*
- Location\_of relation between *rectal* and *adénocarcinome rectal bifocal*
- Localization\_of relation between *bifocal* and *adénocarcinome rectal bifocal*
- Treats relation between *opéré* and *adénocarcinome rectal bifocal*
- Before relation between *opéré* and *Cure de chimiothérapie*
- During relation between *opéré* and *octobre 96*



**Text** Ac folinique 200 mg/m2/j soit 366 mg/j de J1 à J5 en Bolus

**Translation** 'Bolus folinic acid 200 mg/m2, 24 hs, corresponding to 366 mg, 24 hs from day 1 to day 5'.

**Annotations**

- *Ac folinique* as Chemicals\_Drugs
- *200 mg/ms* as Dosage
- */j* as Temporal entity of type Frequency (there are two instances)
- *366 mg* as Dosage
- *J1* as Temporal entity of type Date
- *J5* as Temporal entity of type Date
- *Bolus* as MedicalProcedure
- HasDosage relation between *Ac folinique* and *200 mg/ms*
- HasDosage relation between *Ac folinique* and *366 mg*

- During relation between *Ac folinique* and *Jj* (there are two instances)
- Begins\_on relation between *Ac folinique* and *J1*
- Ends\_on relation between *Ac folinique* and *J5*
- Used\_for relation between *Ac folinique* and *Bolus*

## Discussion of annotation choices / FAQ

### Pronouns

- How do I annotate pronouns?

Pronouns should only be annotated for [Persons](#) entities.

### Temporal expressions

- How should vague temporal mentions such as *récent*, *ancien*, *depuis quelques jours* be annotated?

According to TimeML, these expressions are under-specified dates and durations. They should be marked, but no formal temporal normalization will be assigned to them.

### Assertions

- Which assertion should be used for the mention *présence*? Is Present or Possible suitable?

In general, *présence* should be annotated with the assertion Present.

### Co-reference and temporal relations

- Which relations should be annotated, and which relations should be left for inference from co-reference chains (not currently annotated)?

Relations should only be marked within the same sentence, and between the closest mentions of two entities in case of co-referring entities.

If a co-reference relation seems to apply (typically: identical follow-up mention of a previous entity, specification), the relation will be created during the co-reference annotation process. If still in doubt (two mentions seem to refer to the same concept/object, but there may actually be two different concept/objects), a possibly under-specified - but true - relationship may be used between the two entities, e.g. Simultaneous. Relations between each entity and other entities should then be marked using the proximity rule, and not duplicated.

- What should I do about lexical coreference? It is sometimes ambiguous with temporal relationships such as overlap, or simultaneous.

As explained above, co-reference relations are out of scope for this annotation project. However, temporal relations should be marked. In doubt, temporal relations can be annotated.

- When should temporal relations be annotated? Other (non-temporal) relations sometimes imply a temporal relation. In that case, should it be annotated? (e.g.: X causes Y implies X before Y)

Temporal relations that are explicitly stated in the text should be marked. When a temporal relation can be inferred from another, explicit, relation (e.g.: X causes Y implies X before Y), the temporal relation should *not* be marked. However, if there is a doubt whether the other relation is indeed stated (e.g. X *may* cause Y, it's unclear whether X did cause Y...) whereas the temporal relation is clearly true *only* the temporal relation should be marked.

### Miscellaneous relations

- It is sometimes difficult to create relations with entity of type [Localization](#), because no relation seems to really fit adequately.

[Localization](#) entities and modelization of spatial concepts in general are a limitation of our annotation scheme. The current Localization\_of relation should be divided into several more specific relations. This change may be implemented in future versions of the annotation scheme. Currently, the Localization\_of relation should be used to approximate all spatial relations.

- How should I annotate the mention *suivi des douleurs gastriques*?:

[suivi] des [douleurs gastriques] : Relation Treats between Procedure *suivi* and Disorder *douleurs gastriques*

## 3. BRAT Configuration files

- [annotation\\_conf.pdf](#)
- [tools\\_conf.pdf](#)
- [visual\\_conf.pdf](#)

## References

1. G Savova, W Styler, D Albright, M Palmer, D Harris, G Zaramba, P Haug, C Clark, S Wu, D Ihrke (2012) **SHARP template annotations: Guidelines**. Technical report, Mayo Clinic; web , Retrieved April 22, 2013
2. Aurélie Névéol, Rezarta Islamaj Doğan, Zhiyong Lu (2011) **Semi-automatic semantic annotation of PubMed queries: a study on quality, efficiency, satisfaction**. *Journal of biomedical informatics* 44 (2) pp. 310–318. Elsevier.
3. Olivier Bodenreider, Alexa T McCray (2003) **Exploring semantic groups through visual approaches**. *Journal of Biomedical Informatics* 36 (6) pp. 414–432. Elsevier.
4. Alexa T McCray, Anita Burgun, Olivier Bodenreider (2001) **Aggregating UMLS semantic types for reducing conceptual complexity**. *Studies in health technology and informatics* 84 (0 1) pp. 216. NIH Public Access.