











MANUEL D'ANNOTATION

Ce manuel d'annotation est un manuel participatif. Il vise à aider l'annotation d'entités nommées dans le corpus Democrat "à la façon LitBank".

Les EN retenues et étiquettes

Nos annotations visent à tester et proposer une annotation similaire à celle de David Bamman :

- David Bamman, Sejal Popat and Sheng Shen, "[An Annotated Dataset of Literary Entities](#)," NAACL 2019.
- <https://github.com/dbamman/litbank>

 dbamman added quotations	a371cd6 on 30 Apr 2020
 coref	removed two duplicate mentions and corrected COREF/COP tags
 entities	removed old entity tagger
 events	added event annotations
 img	added coref data
 original	init
 quotations	init
 tagger	Update README.md
 .gitignore	init
 README.md	added quotations

La liste des étiquettes

Les exemples donnés dans les sections *entities* et *events* sont extraits des travaux de David Bamman.

entities

6 étiquettes sont utilisées dans l'annotation des entités nommées dans LitBank (cf. [github Bamman](#)) :

- *People* (PER): *Tom Sawyer, her daughter*
- *Facilities* (FAC) : *the house, the kitchen*
- *Geo-political entities* (GPE): *London, the village*
- *Locations* (LOC) : *the forest, the river*
- *Vehicles* (VEH) : *the ship, the car*
- *Organizations* (ORG) : *the army, the Church*

étiquettes supplémentaires

- HIST

Alors que nous voulions annoter les EVENT - comme dans le LitBank - nous nous sommes trouvés confrontés à un problème.

Dans Democrat, seuls les GN sont annotés. Il nous fallait donc créer une nouvelle référence. Ce faisant, nous avons annotés les événements historiques différemment des EVENT (cf. plus bas dans le manuel).

Sont annotés HIST les groupes nominaux liés à des événements historiques attestés, ex: 10 jours de décembre, la décapitation de Louis XVI

Matt Sims, Jong Ho Park and David Bamman, “Literary Event Detection,” ACL 2019.

- TIME

Sont annotés TIME les références au temps : périodes de la journée (“matin”, “soir”), fêtes religieuses, les durées (“3 heures” = pendant 3 heures), ...

- NO_PER

Pour annoter les animaux, ou autres “caractères” (objets personnifiés, etc) prenant part à l’action.

- METALEPSE

Pour annoter les références au locuteur, narrateur, ...

- TO_DISCUSS : méta étiquette qui permet d’isoler les étiquettes sur lesquelles nous souhaitons revenir

Quelques principes généraux

Le corpus Democrat comprend une annotation de chaînes de coréférence. Toutes les expressions référentielles des textes ont été repérées et regroupées selon leur référence dans des chaînes. Ainsi, les objets pouvant être des entités nommées sont déjà repérées dans le corpus. Pour typer les EN, il nous “suffit” de typer les chaînes, ou de les écarter quand elles ne constituent pas d’entité nommée visée par notre annotation. Quelles sont les stratégies et contraintes d’une telle annotation ?

Lors des réunions de début de projet, nous avons souligné l’idée de conserver au maximum les annotations Democrat, ne pas se conformer au format “*LitBank*” à tout prix. Il ne s’agit pas de dénaturer les annotations Democrat, mais de voir comment et à quel point il est possible d’utiliser l’existant pour donner par la suite ce corpus comme base d’apprentissage pour le pipeline BookNLP.

Manuel d’annotation LitBank :

https://docs.google.com/document/d/1K3q9eYOLiySFq44nUceF38s59Mem_AVdOWSvXToXw8U/edit#

Manuel d’annotation Democrat : https://www.lattice.cnrs.fr/democrat/files/ANR-15-CE38-0008-DEMOCRAT_livvable_methodo.pdf

Ainsi typer les EN dans le corpus Democrat revient à :

- typer l’ensemble de la chaîne. Le typage de l’EN est appliqué à la chaîne entière. Ainsi, annoter une chaîne revient à typer l’ensemble des maillons (sans exception!) de la chaîne avec une seule et même étiquette.

- une chaîne constituée d'un seul maillon (repérée comme singleton dans Democrat : *SI-num*) peut être annotée. En effet, dans LitBank, une chaîne peut être constituée d'un seul maillon.
- Cette façon de travailler a comme inconvénient qu'il faut parfois regarder plus loin que le premier maillon de la chaîne. Par exemple, il est possible qu'une chaîne qui comprend 'La rue Trucmuche' (qui doit recevoir l'étiquette de FAC) commence par le mot 'où'. Dans ce cas, il faut tout de même projeter l'étiquette FAC sur l'ensemble des maillons de la chaîne.

Exemples

Cette section regroupe les exemples issus du corpus Democrat, de nos annotations. Il peut s'agir d'exemples illustratifs/typiques, ou d'exemples problématiques sur lesquels nous avons statué.

People (PER)

By person we describe a single person indicated by a proper name (Tom Sawyer) or common entity (the boy); or set of people, such as her daughters and the Ashburnhams.(Bamman)

Text	Source
sous le règne de Louis Xiii ,	Le Capitaine Fracasse
une de ces gentilhommières si communes en Gascogne, et que les villageois décorent du nom de château	Le Capitaine Fracasse
[...] qui paraissait être le lit de l'unique valet du manoir	Le Capitaine Fracasse
... comme un polisson ...	???
En haut, sur la charge des légumes, allongés à plat ventre, couverts de leur limousine à petites raies noires et grises, les charretiers sommeillaient, les guides aux poignets.	Le Ventre de Paris
Madame François , adossée à une planchette contre ses légumes, regardait, ne voyait rien, dans la maigre lueur jetée à gauche par la petite lanterne carrée, qui n'éclairait guère qu'un des flancs luisants de Balthazar.	Le Ventre de Paris
... exemple...	... titre roman ...

On annote :

- les foules, les collectifs, ... : "la famille", "l'entourage", "les gens d'esprit" ...
- toute personne qui peut entrer dans une pièce : "Dieu"
- ...

On n'annote pas :

- Les parties du corps humain : "les poignets",...

- Les interjections : “Dieu”! qu’il fait beau, ...
- ...
- “cadavre” (chaîne n°586 et n°825 de Pauline) car comparaison

Facilities (FAC)

For our purposes, a facility is defined as a “functional, primarily man-made structure” designed for human habitation (buildings, museums), storage (barns, parking garages), transportation infrastructure (streets, highways), and maintained outdoor spaces (gardens). We treat rooms and closets within a house as the smallest possible facility. (Bamman) > Exemples (complètent la définition ci-dessus):

- *the place* -> bleak house

Règle proposée: on annote comme “facility” ce qui peut être habité (tour, chambre, placard) ou qui peut servir de chemin (sentier, ornière, chaussée).

Text	Source
Deux tours rondes, coiffées de toits en éteignoir , flanquaient les angles d’un bâtiment,	Le Capitaine Fracasse
une de ces gentilhommières si communes en Gascogne , et que les villageois décoient du nom de château.	Le Capitaine Fracasse
une de ces gentilhommières si communes en Gascogne , et que les villageois décoient du nom de château.	Le Capitaine Fracasse
Le chemin qui menait de la route à l’habitation s’était réduit, par l’envahissement de la mousse et des végétations parasites, à un étroit sentier blanc semblable à un galon terni sur un manteau râpé	Le Capitaine Fracasse
Au seuil du chenil , un chien unique, flottant dans sa peau trop large où ses muscles détendus	Le Capitaine Fracasse
...et prit le chemin de la cuisine sans proférer une seule parole.	Pauline
Aussitôt l’étrangère ordonna qu’on mît sa voiture sous la remise et qu’on lui préparât une chambre (ch224)	Pauline
Mais non pas, madame ; vous êtes à Saint-Front, route de Paris , hôtel du Lion-Couronné (chaine183)	Pauline
Par là, du côté des Halles .	Le Ventre de Paris
L’avenue plate s’étendait, avec ses lignes de grands arbres et de maisons basses, ses larges trottoirs grisâtres , tachés de l’ombre des branches, les trous sombres des rues transversales, tout son silence et toutes ses ténèbres	Le Ventre de Paris
C’est à deux pas, rue Montorgueil , au	Le Ventre de Paris

Text	Source
Compas d'or.	
la pointe Saint Eustache	Le Ventre de Paris
... exemple...	... titre roman ...

On annote

Est annoté FAC un espace dans lequel on peut se tenir + pas mobile

- tout passage destiné à la circulation est FAC : “route”, “chemin”, “ornière”, ...
- les foires et marchés sont FAC
- ...

On annote pas

- Les objets/meubles : siège, ‘banc’, ‘table’, etc.
- une partie d'un tout : “les angles”, “la façade”, “la cheminée” (qui fume, sur le toit), “toit”

Geo-political entities (GPE)

Geo-political entities are single units that contain a population, government, physical location, and political boundaries. In literary data, this includes not only cities that have known geographical locations within the real world (London, New York), or nations (England, the United States), but also both named and common imagined entities as well (the town, the village). (Bamman)

Règle : Sont les GPEs les instances gouvernables, contrairement aux LOCs qui sont des lieux sans gouvernement.

Text	Source
Sur le revers d'une de ces collines décharnées [...], s'élevait [...] une de ces gentilhommières si communes en Gascogne	Le Capitaine Fracasse
Sur le revers d'une de ces collines décharnées qui bossuent les Landes [...] s'élevait [...] une de ces gentilhommières	Le Capitaine Fracasse
Un tombereau de choux et un tombereau de pois, au pont de Neuilly , s'étaient joints aux huit voitures de navets et de carottes qui descendaient de Nanterre ; et les chevaux allaient tout seuls, la tête basse, de leur allure continue et paresseuse, que la montée ralentissait encore.	Le Ventre de Paris
Échappé de Cayenne , où les journées de décembre l'avaient jeté, rôdant depuis deux ans dans la Guyane hollandaise , avec l'envie folle du retour et la peur de la police impériale, il avait enfin devant lui la chère grande ville , tant regrettée, tant désirée.	Le Ventre de Paris

Text	Source
... exemple...	... titre roman ...

On annote

- villes, pays, régions, contrées, ...

On n'annote pas

- voir la catégorie LOC et ORG

Locations (LOC)

Locations describe entities with physicality but without political entities. In our dataset, this includes named regions without political organization (New England, the South) and planets (Mars). The most common class, however, are geologically designated areas describing natural settings, such as the sea, the river, the country, the valley, the woods, and the forest. (Bamman)

Exemples (complètent la définition ci-dessus):

- *the side of the hall* -> bleak house
- *up the river , where it flows among green aits and meadows* -> bleak house
- *his dominions* -> the masque of the red death
- *this part of England* -> tess of the durbervilles a pure woman
- *general surroundings* -> bartleby the scrivener a story of wallstreet
- *the Rabbit-Hole* -> alices adventures in wonderland
- *a rat-hole* -> alices adventures in wonderland

Text	Source
Sur le revers d' une de ces collines décharnées qui bossuent les Landes,[..]	Le Capitaine Fracasse
des moellons effrités aux pernicieuses influences de la lune	Le Capitaine Fracasse
Poussez-moi ça dans le ruisseau !	Le Ventre de Paris

On annote

- Astres (soleil, lune) annotés en LOC

On n'annote pas

- le ciel
- un composant / partie d'un tout: Ce qui compose un élément (partie d'un tout) n'est pas EN. Dans le GN **le revers d'une de ces collines décharnées**, **le revers** est une partie de la colline. Il n'est pas annoté. En revanche sont annotées les 2 EN "locations" : "le revers d'**une de ces collines décharnées**" et "le revers d'une de **ces collines décharnées**". Elles réfèrent toutes les deux à un exemplaire d'une classe donnée.

- Les distances et orientations: “sans un seul être vivant à **dix lieues** à la ronde”, “Pierre se tenait debout à **quelque distance** de là, immobile”, “**au-delà** de la grande porte, il y avait...”
- Des endroits insignifiants, trop peu précis pour savoir où c’est exactement : **un tas de carottes, le long du trottoir** (attention dans le dernier exemple **trottoir** est annoté FAC).

Vehicles (VEH)

For our annotations purposes, we define a vehicle as a “physical device primarily designed to move an object from one location to another.” Ships, planes, subway trains, tractors, cars, and bikes would all be considered vehicles (as well as anything else that fits this definition). (Bamman)

Text	Source
Deux ornières remplies d’eau de pluie et habitées par des grenouilles témoignaient qu’anciennement des voitures avaient passé par là	Le Capitaine Fracasse
dernier débris d’ un carrosse défunt sous le règne précédent	Le Capitaine Fracasse
... mais qu’il fallait attendre que l’attelage qui venait de conduire la patache fût un peu rafraîchi.	Pauline
Florent aperçut, au fond d’ une grande calèche , des femmes masquées, les épaules nues, la voix rieuse, se fâchant de ne pouvoir passer, faisant les dégoûtées devant « ces forçats qui n’en finissaient plus. »	Le Ventre de Paris
... exemple...	... titre roman ...

On annote

- chevaux (si non caractère), carosse, voiture, train, etc

On n’annote pas

- Les chevaux qui sont personnifiés (voir NO_PER)

Organizations (ORG)

Organizations are defined by the criterion of formal association. Commonly referenced organizations include the army and the Church (as an administrative entity, distinct from the church as a facility with a physical location). (Bamman)
Exemples (complètent la définition ci-dessus):

- *my business* -> bartleby_the_scrivener_a_story_of_wallstreet

Text	Source
Échappé de Cayenne, où les journées de décembre l’avaient jeté, rôdant	Le Ventre de Paris

Text	Source
depuis deux ans dans la Guyane hollandaise, avec l'envie folle du retour et la peur de la police impériale , il avait enfin devant lui la chère grande ville, tant regrettée, tant désirée.	
... exemple...	... titre roman ...

événements historiques (HIST)

time (TIME)

Attention : cette étiquette n'a pas de correspondance dans le LitBank anglais et a été développée par notre équipe.

Text	Source
sous le règne de Louis XIII ,	Le Capitaine Fracasse
Une nappe de lierre enveloppant à demi l'une des tours tranchait heureusement par son vert sombre sur le ton gris de la pierre déjà vieille à cette époque .	Le Capitaine Fracasse
Il y a trois ans , il arriva à Saint-Front, petite ville ...	Pauline
Moi, je ne pourrais pas rester dans ce diable de Paris, toute la journée , sur un trottoir.	Le Ventre de Paris
Le soir , ils avaient mangé un lapin.	Le Ventre de Paris
Au bout de six semaines , en janvier, un geôlier le réveilla, une nuit, l'enferma dans une cour, avec quatre cents et quelques autres prisonniers.	Le Ventre de Paris

on annotate

- les informations temporelles
- les indications de durée
- les moments de la journée (le jour, la nuit, la matinée, ...)

on annotate pas

- les âges (exemple : jeune fille de seize ans). TIME est une référence qui doit pouvoir être projetée sur une frise chronologique.

NO_PER

Règle : Sont annotés en NO_PER : Les animaux ou objets personnifiés qui ont une importance dans la narration : le chien ou le chat de la famille, ...

Attention : cette étiquette n'a pas de correspondance dans le LitBank anglais et a été développée par notre équipe.

Text	Source
Un vieux chat noir , maigre, pelé comme un manchon hors d'usage et dont le poil tombé laissait voir par places la peau bleuâtre	Le Capitaine Fracasse
une statue de marbre représentant une divinité mythologique, Flore ou Pomone	Le Capitaine Fracasse
Balthazar, le cheval de madame François , une bête trop grasse, tenait la tête de la file.	Le Ventre de Paris

on annote pas

Text	Source
La grande voix des Halles grondait plus haut ; par instants, des volées de cloche, dans un pavillon éloigné, coupaient cette clameur roulante et montante.	Le Ventre de Paris

METALEPSE

Règle : On annote en métalepse les éléments qui réfèrent à des éléments en dehors du récit (métanarration), par exemple au narrateur ou au locuteur.

Attention : cette étiquette n'a pas de correspondance dans le LitBank anglais et a été développée par notre équipe.

Text	Source
Sur la bande frayée à travers les mauvaises herbes, et détrempée par une averse récente, on ne voyait aucune empreinte de pas humain	Le Capitaine Fracasse
Les ronces, aux ergots épineux, se croisaient d'un bord à l'autre des sentiers et vous accrochaient au passage pour vous empêcher d'aller plus loin et vous dérober ce mystère de tristesse et de désolation.	Le Capitaine Fracasse
Comme le lecteur doit être las de cette promenade à travers la solitude	Le Capitaine Fracasse

Cas spécifiques

EN coordonnées

1. Deux entités coordonnées de même type.

exemple 1 :

Text	Source
Sur le revers d'une de ces collines décharnées qui bossuent les Landes, entre Dax et Mont-de-Marsan , s'élevait, sous le règne de Louis Xiii, une de ces gentilhommières si communes en Gascogne	Le capitaine fracasse

Dax et Mont-de-Marsan est annoté comme "GPE". *Dax* et *Mont-de-Marsan* sont, par ailleurs, annotées comme GPE séparément, en tant que têtes de leurs chaînes respectives.

2. Deux entités coordonnées de types différents.

Si deux éléments coordonnés sont de types différents, les éléments sont annotés individuellement mais le segment contenant les 2 éléments coordonnés n'est pas annoté.

Text	Source
[...] mais que les années et l'usage rendaient d'un roux pisseux	Le capitaine fracasse

Dans Democrat, 3 segments repérés : **les années et l'usage**, **les années**, **l'usage**, seul **les années** est annoté TIME.

EN imbriquées

The targets of annotation here include both named entities (e.g., Tom Sawyer) and common entities (the boy). (Bamman, <https://github.com/dbamman/litbank>)

Text	Source
Sur le revers d' une de ces collines décharnées qui bossuent les Landes,[..]	Le Capitaine Fracasse
Sur le revers d'une de ces collines décharnées qui bossuent les Landes,[..]	Le Capitaine Fracasse

Discussion/problèmes

Notre travail est fonction du repérage des unités dans Democrat :

Dans l'exemple suivant, **une des tours** est repéré dans la segmentation Democrat mais pas **(des) tours** (cf. prep+art):

Une nappe de lierre enveloppant à demi l'**une des tours** (FAC)

Alors que l'on aurait pu avoir 2 segments isolés, l'un imbriqué dans l'autre : **une des tours** et **des tours**. C'est ce que nous avons dans l'exemple :

Sur le revers d'**une de ces collines décharnées** qui bossuent les Landes,[...]

2 segments sont repérés, l'un imbriquant l'autre - **ces collines décharnées** (LOC) dans **une de ces collines décharnées** (LOC).

Ressemble à un EN, mais n'a pas de référence

Text	Source
il retournait doucement à l'état de hallier ou de forêt vierge	Le Capitaine Fracasse
Il était comme un fou furieux	Le Ventre de Paris
Quand il déboucha dans la grande rue du milieu, il songea à quelque ville étrange , avec ses quartiers distincts, ses faubourgs, ses villages, ses promenades et ses routes, ses places et ses carrefours , mise tout entière sous un hangar, un jour de pluie, par quelque caprice gigantesque.	Le Ventre de Paris

La métonymie

Ne pas annoter les métonymies ... sauf si la métonymie est le premier maillon d'une chaîne de plusieurs maillons où apparaît l'entité.

Text	Source
Il y a six pavillons, de ce côté-là ; puis de l'autre côté, en face, il y en a encore quatre : la viande, la triperie, la Vallée ...	Le Ventre de Paris
Trois ou quatre , au contraire, au fond de trous d'ombre, semblaient près de tomber sur le nez.	Le Ventre de Paris
une voix appela Claude » encore une métonymie	Le Ventre de Paris

Le pronom "on"

"on" est à annoter de manière différente :

- soit il réfère à une personne ou groupe de personnes. Il est alors annoté "PER"

paul et moi on

- soit il réfère au lecteur, à l'auteur. Il est alors annoté "METALEPSE"

Text	Source
Sur la bande frayée à travers les mauvaises herbes, et détrempée par une averse récente, on ne voyait aucune empreinte de pas humain	Le Capitaine Fracasse

- soit il ne réfère pas, est indéfini. Il n'est pas annoté :

EN "pluriel"

Que faire si l'unité réfère à des éléments de type différents ?

Text	Source
On voyait qu' ils se connaissaient de longue main et se tenaient souvent compagnie dans la solitude du château	Le capitaine fracasse

ils et *se tenaient* réfère au chien et son maître (pers + NoPers) (...nb : en outre, il est possible que nous fusionnons par la suite pers et nopers)

Ne sont pas EN (ie le TYPE_REFERENT reste non renseigné)

Nous avons discuté et écarté les annotations suivantes :

- *body parts* -> Parties du corps
- *character* -> Au lieu de 'personne', car il y a des *caractères* (au sens de personnage) qui ne soient pas une personne
- *environment/ambiance* -> "dans le noir", des descriptions de l'environnement et de l'ambiance qui ne soient pas tout à fait temporelles ou locatives
- *artifact** -> Des objets fabriqués par des humains qui ne soient pas de facilities

à noter : la plupart des éléments désignés ici pourrait être l'objet de repérage automatique dans le texte, à partir de lexiques donnés (animaux, nourriture, partie du corps, ...).

AINSI NE SONT PAS A ANNOTER les exemples/catégories suivantes :

- Les animaux qui appartiennent au décor (? à mieux définir)

Text	Source
Deux ornières remplies d'eau de pluie et habitées par des grenouilles témoignaient qu'anciennement des voitures avaient passé par là	Le Capitaine Fracasse

- nommage du monde (description), entités non finies

Text	Source
Une nappe de lierre enveloppant à demi l'une des tours tranchait heureusement par son vert sombre sur le ton gris de la pierre	Le Capitaine Fracasse
au-dessus des genêts et des bruyères ,	Le Capitaine Fracasse

- éléments de la nature

Text	Source
Deux ornières remplies d' eau de pluie et habitées par des grenouilles	Le Capitaine Fracasse
Sur la bande frayée à travers les mauvaises herbes, et détrempée par une averse récente , on ne voyait aucune empreinte de pas humain	Le Capitaine Fracasse

- notion et termes abstraits

Text	Source
mais, en approchant, son avis se fût modifié	Le Capitaine Fracasse
par l’envahissement de la mousse et des végétations parasites	Le Capitaine Fracasse
mais la sécurité de ces batraciens montrait une longue possession et la certitude de n’être pas dérangés	Le Capitaine Fracasse
Il retomba dans le bien-être égoïste de sa condition , fit le gros dos (chaine87)	Pauline
Mille images confuses (ch 93) passaient dans ses rêves (ch94) et la réveillaient en sursaut	Pauline

- habits, mobilier, ...

Text	Source
semblable à un galon terni sur un manteau râpé	Le Capitaine Fracasse
et voilaient à demi une statue de marbre représentant une divinité mythologique	Le Capitaine Fracasse

- objets et oeuvres fabriquées de la main de l’homme

Text	Source
Entre les Hercules, dans des niches peintes, se pavanaient des bustes d’empereurs romains et autres personnages illustres de l’histoire ;	Le Capitaine Fracasse

- partie du corps

Text	Source
une jeune et jolie femme de chambre qui avançait à la portière sa tête entourée de foulards en désordre	Pauline

- Référence (+/- abstraite?) à un groupe social, un ensemble de personnes, ...

Text	Source
Qu’eût-on dit, en effet, de voir le baron de Sigognac accoutré comme un gueux de l’Hostière ou comme un cueilleur de pommes du Perche	Le Capitaine Fracasse
Aussi beaucoup de gens croyaient-ils que les Sigognac étaient éteints,	Le Capitaine Fracasse

- ...

Annotation des EVENT

Chez Bamman :+1:

- *My father's eyes had **closed** upon the light of this world six months, when mine **opened** on it.* -> Dickens, David Copperfield
- *Call me Ishmael.* -> Melville, Moby Dick
- *His sister was a tall, strong girl, and she **walked** rapidly and resolutely, as if she knew exactly where she was going and what she was going to do next* -> Cather, O Pioneers

Pour voir comment c'est fait dans Litbank :

https://people.ischool.berkeley.edu/~dbamman/pubs/pdf/acl2019_literary_events.pdf

Quelques exemples au format BRAT:

https://apps.lattice.cnrs.fr/brat/index.xhtml#/litbank/events/1064_the_masque_of_the_red_death_brat

https://apps.lattice.cnrs.fr/brat/index.xhtml#/litbank/events/1023_bleak_house_brat

Un article de Baman où il y a quelques explications sur l'annotation des EVENTS:

<https://aclanthology.org/P19-1353.pdf>

Et aussi le readme de Litbank:

<https://github.com/dbamman/litbank/blob/master/README.md>

Nos règles d'annotation

Règle : Les events peuvent être : des noms, des verbes et des adjectifs.

Un verbe à l'infinitif peut être annoté :

exemple : il entendit le jeune homme **entrer**

Règle : Les events ont les catégories sémantiques suivantes : les états, changements d'état explicites, activités, des accomplissements, les réalisations (achievements and accomplishments).

Règle : S'il y a un doute, il faut annoter au plus large, on peut faire des corrections automatiques après.

Règle : dans un groupe verbal, il faut annoter un seul élément - ce peut être le GV dans son ensemble si c'est une expression figée, ou si il s'agit d'une phase (ex: **commencer un repas**), ce peut être uniquement le verbe si c'est le verbe qui est significatif (ex: **improviser** un repas, où improviser signifie préparer, le repas n'a donc pas eu lieu)

Règle : on annote les event niés, quand il s'agit d'une action - sans inclure la négation dans l'annotation (ex: il n'**entendit** pas le jeune homme **entrer**, il ne **retint pas ses larmes**)

Règle : on annote indépendamment du niveau de récit - un récit imbiqué dans un autre est annoté

Text	Source
The "Queen of the Waves" avait fait naufrage	Némoville

Text	Source
Et pour le moment, les naufragés durent céder à une préoccupation plus impérieuse : celle de se reconforter par un peu de nourriture...	Némoville
Sans même prendre la peine d'allumer un feu, chacun improvisa son repas de quelques conserves froides, puis se roula dans sa couverture et s'endormit	Némoville
Deux jeunes gens, Roger de Ville et Paul Lamontagne, offrirent d' aller à la découverte	Némoville
Elle s'adoucit .	Pauline
... lorsque Roger la regardait , rougissante et un peu timide	Némoville
Veuillez me suivre monsieur l'abbé, je vais vous conduire auprès d'un mourant .	Némoville
D'un coup de talon sur le monstre, il revint à la surface.	Némoville

On annote

- Pour les verbes non modaux et non auxiliaires, qui prennent un complément nominal événementiel ou un infinitif (e.g. *commencer, finir, débiter, improviser*), on annote ce verbe et le complément nominal/ infinitif.
- On annote les changements d'état.
- Les noms suivants sont annotés en tant qu'événement: ceux qui désignent la météo (*la pluie, le brouillard, etc*), les noms dérivés de verbes d'action (*crie, exclamation, saut, ...*), les noms prädicatifs (*présence, naissance, mort, introduction; sont souvent aussi des dérivés verbaux*) et des noms désignant des événements (*repas, paroles, etc*).
- On annote les réflexifs avec les verbes.
- On annote les verbes de perception - entendre, se souvenir, ...

On n'annote pas

- compléments directs et indirects des verbes, sauf s'il s'agit d'une expression figée. On annote également une préposition si elle fait partie du verbe (e.g. *donner contre* qui est synonyme du verbe *échouer*).
- Ne sont pas annotés en EVENT : les auxiliaires, mais uniquement le verbe tête.
- Pour les verbes comme *dire, répondre, proposer de*, etc + complément verbal à l'infinitif, on n'annote que les verbes qui désignent l'action de dire/parler, et non pas la chose dite.
- Les 'en' dans des constructions avec des gérondifs ('en voyant', 'en disant', ...) on annote uniquement le participe.

Exemples d'annotations de cas limites, cf le Hackmd Étapes de travail (in progress en ce mardi 28.9)

Text	Source
Pécuchet ne put s'empêcher de dire :	Bouvard et Pécuchet

Comparatif des annotations LATTICE et Democrat

Democrat	Lattice	Source
Jeune fille de (seize ans) TIME	jeune fille de seize ans	Sarrasine