Consignes d’annotation pour la campagne ESV

Version - 1.2

14 mars 2023

**Auteurs**

Claire Nédellec, Elisa Lubrini, Louise Deléger, Marie Grosdidier,

Robert Bossy, Sandy Dupérier, Clara Sauvion

**Destinataires**

Experts annotateurs impliqués dans la campagne d’annotation

| Sommaire  [**1. Introduction 2**](#_heading=h.2eklun59yc2r)  [1.1. Contenu du document 2](#_heading=h.7qf0o050iery)  [**2. Annotation des mentions d’entités 2**](#_heading=h.kbxthzhapg1n)  [2.1. Listes des types d’entité 2](#_heading=h.s4yrao13lppg)  [2.2. Consignes générales d’annotation des entités 2](#_heading=h.2agktncmfe6u)  [2.2.A. Bornes 2](#_heading=h.6f6dvbyckldu)  [2.2.B. Enumération 4](#_heading=h.miqhixtcvj5a)  [2.2.C. Discontinuité 4](#_heading=h.k2r6c9hfqp12)  [2.2.D. Acronymes 5](#_heading=h.xgkhiy1y8xm5)  [2.3. Normalisation 5](#_heading=h.x6ugv639uveo)  [2.3.A. Normalisation des organismes 5](#_heading=h.goc5arhrtzwj)  [2.3.B. Normalisation des voies de dispersion 5](#_heading=h.dj64hxw3dvv1)  [2.3.C. Lieu 6](#_heading=h.ql9jx62si8dn)  [2.4. Consignes d’annotation par type d’entité 6](#_heading=h.xvir7l)  [2.4.A. Organisme nuisible 6](#_heading=h.jaozzg21zfzb)  [2.4.B. Plante hôte 7](#_heading=h.161jxze2bus7)  [Ce type n’inclut pas les plantes désignées par leur usage ou par leur filière : ornemental, cultivated, garden plant, etc. 8](#_heading=h.uchwuoqffxc)  [2.4.C. Vecteur 9](#_heading=h.78snr87o51oj)  [2.4.D. Maladie 10](#_heading=h.qjhyl6w3goae)  [2.4.E. Voie de dispersion 11](#_heading=h.z421nqedf6na)  [2.4.F. Lieu 12](#_heading=h.u6en7v8g346s)  [2.4.G. Date 13](#_heading=h.qv5vukrw0bu5)  [2.4.H. Quantité 14](#_heading=h.v5y1xup7y2lw)  [**3. Annotation des relations 14**](#_heading=h.bal0h9i4xjqg)  [3.1. Relation de Coréférence 14](#_heading=h.s4h5h1euvftt)  [3.2. Relations binaires 15](#_heading=h.8xs82148nvir)  [3.2.A. Cause 15](#_heading=h.o3ikajljrcq8)  [3.2.B. Trouve\_sur 15](#_heading=h.d5hepntr95yw)  [3.2.C. Vecte\_par 16](#_heading=h.p3kyqzdvlp5o)  [3.2.D. S\_exprime\_sur 16](#_heading=h.88if4y2yh6h3)  [3.2.E. Disperse\_par 16](#_heading=h.xbxjbls95c1k)  [3.2.F. Localise\_en 17](#_heading=h.3urh33owwx73)  [3.2.G. Detecte\_le 17](#_heading=h.a061ne5aojwr)  [3.2.H. Quantifie 18](#_heading=h.9wch9pvod3x0)  [3.3. Relations n-aires 18](#_heading=h.eqk1iofmr7t8)  [3.4. Modalité 18](#_heading=h.eatjvzwct149)  [3.4.A. Négation 19](#_heading=h.1tjxpdoiqg1b)  [3.4.B. Hypothèse 19](#_heading=h.kg2eloklcs2g) |
| --- |

# ***Introduction***

## Contenu du document

Ce document détaille les consignes d’annotation manuelle du corpus de documents de la campagne d’annotation ESV. Il est destiné aux experts qui annotent ces documents et à l’évaluation et la mise au point les méthodes qui prédisent automatiquement ces annotations. Il est divisé en deux parties dédiées à l’annotation des entités et à l’annotation des relations. Chaque instruction est détaillée et illustrée par des exemples.

# ***Annotation des mentions d’entités***

## Listes des types d’entité

Les couleurs utilisées dans l'éditeur d'annotation (AlvisAE) sont reprises ici par type d’entité. Dans ce qui suit, nous utilisons ce code de couleurs pour mettre en évidence les mentions d'entités et faire référence aux entités annotées.

| Organisme\_nuisible | Voie\_de\_dispersion | Lieu |
| --- | --- | --- |
| Vecteur | Maladie | Date |
| Plante |  | Quantité |

## Consignes générales d’annotation des entités

### **Bornes**

Les annotations doivent s’étendre à tous les éléments pertinents, mais exclure ceux qui n’apportent pas d’informations pour identifier l’entité. Notamment, les mots qui n’ont pas d’intérêt hors contexte ne sont pas annotés. Les annotations sont restreintes à un groupe nominal. Elles incluent éventuellement une proposition relative si elle contribue à la description de l’entité.

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  | Adjectif inutile | Field samples of wilting chickpea plants with chlorotic and  necrotic roots were taken to the laboratory at CENSA. | Field samples of wilting chickpea plants with chlorotic and  necrotic roots were taken to the laboratory at CENSA. |
|  | Entités distinctes | In Bid To Boost Exports, Algeria Plants Millions Of Olive Trees | In Bid To Boost Exports, Algeria Plants Millions Of Olive Trees |
| (3) | Adjectifs utiles | Gram-negative bacterium Candidatus Liberibacter species | Gram-negative bacterium Candidatus Liberibacter species |

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  | inclure ce qui aide à identifier | PWD occurs when the host plant is infected with pinewood nematodes. | FPWD occurs when the host plant is infected with pinewood nematodes. |
|  | exclure ce qui n’aide pas à identifier | the existing mosquitocidal B. thuringiensis strains | the existing mosquitocidal B. thuringiensis strains |

Les noms des types d’entité (insect, pest, …) sont inclus dans l’annotation lorsqu’ils qualifient une entité. Dans les exemples (5) et (6), les entités sont spécifiées par les noms des espèces et les types des entités. Dans les exemples, exemples (5) et (7) les adjectifs précisent la nature de l’entité.

| Exemples | | | |
| --- | --- | --- | --- |
| inclure types d’ON | | **Correct** | **Incorrect** |
|  | nom | *Bactrocera dorsalis* insect | *Bactrocera dorsalis* insect |
|  | nom | fall armyworm pest | fall armyworm pest |
|  | adjectif | Basque Country | Basque Country |

À remarquer la différence entre adjectifs qui ajoutent de l’information pour spécifier l’entité (e.g. (5) Bactrocera dorsalis insect) et les adjectifs qui décrivent l’entité sans donner plus d’information sur son identité (e.g. (1) wilting chickpea).

Les attributs qui font partie du nom de l’entité sont annotés comme dans l’exemple (8).

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  | partie du nom (annoté) | yellow mottle virus | yellow mottle virus |
|  | description (pas annoté) | found on the steep Mont Blanc massif | found on the steep Mont Blanc massif |

### **Enumération**

Si les éléments énumérés sont tous pertinents indépendamment les uns des autres, ils sont annotés séparément.

Si une énumération de termes qualifie la même entité, les termes sont annotés séparément. Si l’énumération est composée de termes de plus en plus spécifiques ou généraux, seule l’entité la plus spécifique participe aux relations.

| Exemples | | | | |
| --- | --- | --- | --- | --- |
|  | |  | **Correct** | **Incorrect** |
|  | région, pays |  | During the 2014 - 2015 harvest, field samples of wilting chickpea plants were taken to the laboratory at the CENSA, Mayabeque Province , Cuba. | During the 2014 - 2015 harvest, field samples of wilting chickpea plants were taken to the laboratory at the CENSA, Mayabeque Province , Cuba. |

Si pour une énumération données, seul un sous-ensemble des éléments énuméré est pertinent, sans qu’on sache quel sou-ensemble alors l'énumération est annotée comme une seule entité.

Les exemples (11) à (13) illustrent ces cas. Dans l’exemple (13), une bactérie est isolée dans une région qui comprend plusieurs territoires énumérés, mais elle n’est pas nécessairement présente dans chacun d’entre eux.

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  | séparés | The bacteria can be found in both Italy and France | The bacteria can be found in both Italy and France |
|  | séparés | The bacteria can be found in either Italy or France | The bacteria can be found in either Italy or France |
|  | ensemble | The bacteria was found in an ancient administrative region that now ecompasses territories across Italy and France | The bacteria was found in an ancient administrative region that now ecompasses territories across Italy and France |

### **Discontinuité**

Les annotations d’entités peuvent être discontinues, en particulier quand un même adjectif ou un même nom qualifient deux entités distinctes. Dans l’exemple (14) Europe qualifie à la fois Nothern et Southern qui sont deux régions distinctes.

| Exemples | | | | |
| --- | --- | --- | --- | --- |
|  | |  | **Correct** | **Incorrect** |
|  | 2 entités:  Northern Europe  et  Southern Europe |  | | Northern | and Southern.. | Europe | | --- | --- | --- | | Northern and Southern Europe |
|  | Northern and Southern Europe |

### **Acronymes**

Les acronymes suivant une mention d’entité et qui désignent la même entité sont annotés séparément. Ils sont connectés par une relation de coréférence, voir la [section correspondante](#_heading=h.wnyagw).

| Exemples | |
| --- | --- |
|  | Fusarium oxysporum f. sp. Cubense fungi (Foc) |
|  | Huanglongbing (HLB) |

## Normalisation

### **Normalisation des organismes**

Les entités de type organisme (organismes nuisibles, vecteurs et plantes) sont associées à leur taxon le plus précis possible, désigné par son identifiant taxonomique.

Le référentiel choisi est la taxonomie du NCBI dont le navigateur est à : .<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>

Dans le cas où l’organisme n’est pas dans la taxinomie du NCBI, il devra être annoté par l’identifiant de Enclyclopedia of Life (EOL) : <https://eol.org/>

Dans le cas où l’organisme n’est dans aucun des deux référentiels, il est normalisé par la classe de l’ontologie OntoBiotope, désignée par son identifiant.

#### Exemples

*Xylella* peut apparaître avec juste le nom du taxon mais en fonction du contexte du document **il faut** noter le numéro NCBI de l’espèce comme par exemple *Xylella fastidiosa.*

### **Normalisation des voies de dispersion**

Les entités de type voie de dispersion sont normalisées par une classe de l’ontologie OntoBiotope.

### **Lieu**

Les lieux sont normalisés par l’identifiant donné par le référentiel GeoNames : <https://www.geonames.org/>

Quand plusieurs entrées sont possibles, par défaut la région administrative est choisie. Il faut soigneusement choisir le bon niveau quand plusieurs entrées sont possibles. Quand le texte est vague et que le lieu administratif est trop restrictif, la région, l’île, le pays peuvent être sélectionnés.

Lorqu’un lieu n’existe pas dans Geonames, il est annoté par l’identifiant de la région englobante la plus spécifique.

## Consignes d’annotation par type d’entité

### **Organisme nuisible**

#### Définition

Un organisme nuisible est un **organisme vivant** ou un groupe d’organismes, dont **la nuisance est explicite dans le document**. Ce sont des ravageurs ou agents pathogènes, des genres bactéries, virus, insectes, champignons, oomycètes, nématodes, gastéropodes ou arachnides.

Ils sont dénotés par leur nom scientifique ou leur nom vernaculaire (entier ou sous forme d’acronyme).

Les mentions d’organismes **non nuisibles** ne doivent pas être annotées.

Lorsque la mention désigne un organisme nuisible par son taxon, seules les mentions de taxon inférieur ou égal au genre sont annotées, c’est-à-dire, les genres, les espèces et les sous-espèces (e.g. souches).

Les mentions d’organsimes nuisibles qui ne correspondent pas à des taxons mais qui sont suffisamment précises pour être informatives sont néanmoins annotées.

Les entités de type organisme nuisible annotées doivent être normalisées (voir [section sur la normalisation](#_heading=h.34g0dwd)).

| Exemples | | |
| --- | --- | --- |
|  | nom scientifique | The *Spodoptera frugiperda* is a species in the order Lepidoptera. |
|  | nom vernaculaire | The name “fall armyworm” can refer to several species. |
|  | nom vernaculaire (acronyme) | An urgent response to the rapid spread of FAW is needed. |

#### Bornes

La mention organisme nuisible inclut tous les qualificatifs utiles à sa détermination et exclut les autres. Notamment, le nom scientifique inclut les mots désignant la **sous-espèce**, **l’autorité** et la **date** le cas échéant.

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  | sous-espèce | *X. fastidiosa subspecies pauca ST53* | *X. fastidiosa subspecies pauca ST53* |
|  | l’autorité | Spodoptera frugiperda (JE Smith) | Spodoptera frugiperda (JE Smith) |
|  | date | Spodoptera frugiperda, Smith, 1797 | Spodoptera frugiperda, Smith, 1797 |

Les organismes nuisibles désignés par leur stade de développement ne sont pas annotés.

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  | trop général pour être une entité | The larvae develop in 10 months | The larvae develop in 10 months |
|  | informations non nécessaire à la détermination de l’entité | adult Popillia japonica | adult Popillia japonica |

#### Exceptions

Certaines mentions d’organsime nuisible sont très générales. Certaines de ces mentions sont tellement fréquentes qu’elles n’apportent aucune information.

Les mentions suivantes sont dans ce cas. Elles ne sont pas annotées si elles sont employées **seules, sans autre précision**, c'est-à-dire sans complément ni adjectif.

* pest
* specimens
* biotope
* carrier
* host
* microbe
* microbial
* microorganism
* population
* subject
* vector
* virus
* snail
* nematode
* phytoplasm
* pathogen
* parasite

### **Plante hôte**

#### Définition

Une plante hôte est une plante ou un groupe de plantes, dont le rôle d’hôte est explicite dans le document. Ce rôle d’hôte peut être représenté par **une relation avec un organisme nuisible** ou **par des termes caractérisant son rôle**.

Le rôle d’hôte doit être compris ici au sens large **d’organisme hôte qui est habitat d’un organisme nuisible,** il inclut les plantes de repos, les plantes réservoirs, et les plantes dans lesquelles l’organisme nuisible se reproduit.

Les entités annotées doivent être normalisées (voir [section sur la normalisation](#_heading=h.34g0dwd)).

| Exemples | | |
| --- | --- | --- |
|  |  | [...] shared by two hosts, *Crepis foetida* and *Daucus carota*. |

Les espèces et genres peuvent être dénotés par leur **nom scientifique** ou leur **nom vernaculaire**.

| Exemples | | |
| --- | --- | --- |
|  |  | Pitch pine (*Pinus rigida* Mill.) |

### **Ce type n’inclut pas les plantes désignées par leur usage ou par leur filière : *ornemental*, *cultivated*, *garden plant*, etc.**

Il n’inclut pas les parties de plantes, par exemple les fruits, ou les tiges.

Les noms de plante qui sont inclus dans des noms de maladie ne sont pas annotés. Par exemple, dans “Tomato brown rot virus”, "Tomato" ne doit pas être annoté comme une plante.

| Contre-exemples | | |
| --- | --- | --- |
|  |  | common plant species such as olive, blackberry, plum, avocado, citrus, cypress and many other woody species”. “as major pest of many fruit crops”, “the oriental fruit fly is known to target over 230 different types of products, including pome, stone fruits, citrus, dates, avocados, peppers and tomatoes |

Lorsque la mention désigne une plante par son taxon, seules les mentions inférieures ou égales au genre sont annotées, c’est-à-dire, les genres, les espèces et les sous-espèces (e.g. cultivar).

Les mentions de plantes qui ne correspondent pas à des taxons mais qui sont suffisamment précises pour être informatives sont néanmoins annotées, par exemple, *salad*.

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  | Asplenium sp. CV 961 | 🏷️ NCBI ID: 218655  (Asplenium sp. CV 961) | 🏷️ NCBI ID: 32071  (Asplenium) |

#### Bornes

La mention Plante hôte inclut tous les qualificatifs utiles à sa détermination et exclut les autres. Notamment, le nom scientifique inclut les mots désignant la sous-espèce, l’autorité et la date le cas échéant.

| Exemples | | |
| --- | --- | --- |
|  |  | *Plectranthus hadiensis var. tomentosus* plants with stem and root rot symptoms |
|  |  | A sentinel plantation of *Prunus domestica* cv. Opal, *Quercus petraea* and *Salix alba* were established . |
|  |  | *Pinus koraiensis* Siebold & Zucc |

#### Exceptions

Certaines mentions désignent une entité très générale. Ces mentions sont tellement fréquentes que, la plupart des fois, elles n’apportent aucune information. Les mentions suivantes ne sont pas annotées si elles sont employées **seules, sans autre précision**, c’est à dire sans complément ni adjectif :

* host
* crop

### **Vecteur**

#### Définition

Un vecteur est un organisme ou un groupe d’organismes, dont le rôle de vecteur ou de transporteur d’organisme nuisible ou de maladie à une plante est explicite dans le document. Ce rôle peut être représenté **par une relation** avec un organisme nuisible ou une plante, ou une maladie **ou par des termes** caractérisant son rôle, par exemple le mot *vector*.

Il peut s’agir d’insectes, d’arachnides, de nématodes de champignons et potentiellement d’autres organismes.

Le rôle de vecteur doit être compris ici au sens large, il inclut les actions de transport externe (ex. poil du thorax) ou interne (salive, pièces buccales). La capacité de transmission d’un organisme nuisible à une plante donnée n’est pas nécessairement mentionnée.

Lorsque la mention désigne le vecteur par son taxon, seules les mentions inférieures ou égales au genre sont annotées, c’est-à-dire, les genres, les espèces et les sous-espèces.

Les mentions de vecteurs qui ne correspondent pas à des taxa mais qui sont suffisamment précises pour être informatives sont néanmoins annotées, par exemple, *sap-sucking insect*.

Les entités de vecteur annotées doivent être normalisées (voir [section sur la normalisation](#_heading=h.34g0dwd)).

| Exemples | | |
| --- | --- | --- |
|  |  | ACP vectors the lethal bacterium (*Candidatus* Liberibacter asiaticus) that causes Huanglongbing |
|  |  | beech bark disease spread by a sap-sucking insect |
|  |  | The African citrus psyllid, *Trioza erytreae* (Hemiptera : Triozidae) is a vector of citrus greening disease (Huanglongbing - HLB) caused by the bacterium *Candidatus* *liberibacter* |

Les espèces et genre de vecteurs sont dénotés par leur **nom scientifique** ou leur **nom vernaculaire**.

| Exemples | | |
| --- | --- | --- |
|  |  | Asian citrus psyllid |
|  |  | T. erytreae, Diaphorina citri |

#### Bornes

La mention de Vecteur inclut tous les qualificatifs utiles à sa détermination et exclut les autres. Notamment, le nom scientifique inclut les mots désignant la sous-espèce, l’autorité et la date le cas échéant.

| Exemples | | |
| --- | --- | --- |
|  |  | *Monochamus alternatus* Hope and *M. saltuarius* (Gebler) (Coleoptera : Cerambycidae) transmit *Bursaphelenchus xylophilus* causing pine wilt disease |

#### Exceptions

Certaines mentions désignent une entité très générique. Ces mentions sont tellement fréquentes que généralement elles n’apportent aucune information. Les mentions suivantes ne sont pas annotées si elles sont employées **seules, sans autre précision**, c’est-à-dire sans complément ni adjectif:

* vector
* insect
* insect vector

### **Maladie**

#### Définition

Une maladie peut être la conséquence de la présence d'un organisme nuisible sur/dans une plante hôte. Elle est désignée par un terme reconnu. Le type d’entité Maladie exclut les symptômes ou signes seuls.

Les termes désignant une plante ou une partie de plante, suivis du mot “disease” ne sont pas annotés s’ils ne représentent pas le nom habituel donné à la maladie, par exemple, *apple tree disease* ne doit pas être annoté.

#### Exceptions

Certaines exceptions de maladie à annoter sont listées ici

* *banana disease*

| Exemples | | |
| --- | --- | --- |
|  |  | Huanglongbing, citrus greening disease, pine wilt disease, beech bark disease, Olive Quick Decline Syndrome (OQDS), almond leaf scorch disease (ALSD) |
|  |  | Common Florida Citrus Disease |

#### Bornes

La mention de Maladie est restreinte aux seuls mots qui la désigne. Les qualificatifs qui représentent la gravité ou l’étendue de la maladie ne sont pas inclus dans l’annotation. Le nom de la plante est inclus dans l‘annotation de la maladie s’il fait partie du nom.

| | Exemples | | | | | --- | --- | --- | --- | |  | | **Correct** | **Incorrect** | |  | gravité | deadly pine wilt disease | deadly pine wilt disease | |  | étendue | widespread pine wilt disease | widespread pine wilt disease | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  | nom de la plante  fait partie du nom de la maladie | pine wilt disease | pine wilt disease |

#### Exceptions

Les noms de maladie qui sont inclus dans le nom de l’organisme ne sont pas annotés. Par exemple, dans la phrase "tomato brown rot virus has been found in a tomato farm", "tomato brown rot virus" est annoté comme un organisme nuisible. "tomato brown rot" ne doit pas être annoté.

Certaines mentions désignent une entité très générique. Ces mentions sont tellement fréquentes que, la plupart des fois, elles n’apportent aucune information. Les mentions suivantes ne sont pas annotées si elles sont employées **seules, sans autre précision**, c’est à dire sans complément ni adjectif :

* disease
* illness

### **Voie de dispersion**

#### Définition

Une voie de dispersion est **abiotique** ou **biotique**. Son rôle est explicite dans le document. Ce rôle peut être représenté **par une relation** avec l’organisme nuisible ou la maladie transmise ou la plante hôte, ou **par des termes** caractérisant son rôle, par exemple le mot *pathway*. Ce sont généralement des parties de plantes (ex. graines, fruits) transformées ou non (ex. palettes de transport en bois), ou plus largement des voies abiotiques, aériennes, maritimes, aquatiques, routières, *etc*.

La voie de dispersion qui permet à un organisme nuisible d’affecter une plante n’est pas nécessairement mentionnée dans le document.

| | Exemples | | | | | --- | --- | --- | --- | |  | Imports are vital to Florida’s economy and can include hitchhikers within the cargo or on/in the vessel (e.g. woodboring beetles in wooden pallets, organisms in ship ballast water). | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

#### Bornes

La mention de voie de dispersion inclut tous les qualificatifs utiles à sa détermination et exclut les autres.

| Exemples | | | |
| --- | --- | --- | --- |
|  | The most common pathway for these pests to enter the State is by “hitchhiking” in fruits and vegetables brought back inadvertently by travelers as they return. | | |

#### Exceptions

Certaines mentions désignent une entité très générique. Ces mentions sont tellement fréquentes que, la plupart des fois, elles n’apportent aucune information. Les mentions suivantes ne sont pas annotées si elles sont employées **seules, sans autre précision**, c’est à dire sans complément ni adjectif :

* transport
* pathway

#### Normalisation

Les entités de type voie de dispersion sont normalisées. (voir [section sur la normalisation](#_heading=h.34g0dwd)).

### **Lieu**

#### Définition

Les entités de type lieu sont des lieux géographiques. Ils sont annotés comme des lieux quand ils font partie d’une autre entité à la condition que que le lieu mentionné soit bien le lieu physique. Les lieux peuvent être désignés par des adjectifs, par exemple *Apulian region*. Certains noms d'organisations, de journaux ou d'instituts comprennent des noms de lieux géographiques. Ces noms de lieux géographiques ne doivent pas être annotés.

Les entités annotées doivent être normalisées (voir [section sur la normalisation](#_heading=h.34g0dwd)).

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  |  | Panama disease  Lissachatina fulica is also known as the   | giant | African | snail | | --- | --- | --- | | Panama disease  Lissachatina fulica is also known as the  giant African snail |

#### Bornes

Les annotations incluent toutes les précisions utiles à l’identification du lieu. Par exemple city of Zürich

#### Exceptions

Les mentions suivantes sont trop générales pour être utiles et ne sont pas annotées :

* worldwide
* all southern countries

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  |  | in many countries, such as France and Spain. | in many countries, such as France and Spain. |
|  |  | The vector is found worldwide. | The vector is found worldwide. |

### **Date**

#### Définition

| Une date est une mention temporelle qui peut être une date entière ou partielle (année-mois-jour ou juste année) ou une période. Elles sont absolues ou relatives.   | Exemples | | | | | --- | --- | --- | --- | |  | | **Correct** | **Incorrect** | |  | date  d’observation | In 1948, the pine wilt disease was found on 3 pine trees in the region. | Spodoptera frugiperda, Smith, 1797 | |  | Période relative | This last decade, the disease spread |  | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

#### Bornes

L'annotation doit inclure la date la plus précise (année, mois/saison, etc).

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  | date  d’observation | In March 1948, the pine wilt disease was found on 3 pine trees in the region. | In March 1948, the pine wilt disease was found on 3 pine trees in the region. |
|  | date  relative | In March 1948, the disease was discovered. The following year its presence was recorded in 5 different European countries. | In March 1948, the disease was discovered. The following year its presence was recorded in 5 different European countries. |

Les intervalles temporels sont annotés comme une seule entité.

| Exemples | | | |
| --- | --- | --- | --- |
|  | | **Correct** | **Incorrect** |
|  | période | In the period January 2017 - March 2020, 60480 news items on 494 pests were selected. | In the period January 2017 - March 2020, 60480 news items on 494 pests were selected. |

#### Exceptions

Les dates mentionnées dans les citations bibliographiques (que ce soit dans le corps du texte ou la section référence) et dans les noms d’espèces ne sont pas annotées.

### **Quantité**

#### Définition

| Une quantité est une valeur numérique absolue, un terme négatif ou encore un pourcentage qui **qualifie** une entité. Une quantité est toujours en relation binaire ou n-aire.   | Exemples | | | | | | --- | --- | --- | --- | --- | |  | |  | **Correct** | **Incorrect** | |  | Quantité |  | In 1948, the pine wilt disease was found on 3 pine trees in the region. | Spodoptera frugiperda, Smith, 1797 | |  | Période relative |  | In Bid To Boost Exports, Algeria Plants Millions Of Olive Trees | losses that can reach 100%. | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

# ***Annotation des relations***

Les relations sont annotées dans le texte seulement lorsque ces dernières sont explicitées dans le texte. Les entités qui sont reliées par une relation sont appelées les arguments de la relation. Les relations sont dites binaires lorsqu’elles relient deux arguments. Elles sont dites ternaires pour trois argument, quaternaire pour quatre. On appelle plus généralement n-aires les relations qui ont plus de deux arguments.

#### Distance

Une entité donnée est reliée à d’autres entités si elles sont distantes au maximum de deux phrases avant et deux phrases après la phrase de l’entité. Les entités plus lointaines ne sont pas reliées magrié leur intérêt, sauf si elles font partie du titre.

#### Type de relation

Il y a deux types de relations, (1) les relations d’équivalence qui relient deux entités qui signifient la même chose. Elles sont appelées relation de coréférence. (2) Les relations thématiques relient des entités en fonction de leur rôle dans l’interaction.

## Relation de Coréférence

La relation de coréférence relie des termes, deux ou plus, qui font référence à la même entité. Elle représente une équivalence entre plusieurs mentions.

| Exemples | | |
| --- | --- | --- |
|  | coréférence entre maladies | Panama disease, also called banana wilt, a devastating disease of bananas |
|  | coréférence entre ONs | Apple proliferation phytoplasma (APP) (Candidatus Phytoplasma mali) is a plant pest |

#### Simplification de l’annotation grâce aux coréférences

Lorsque deux entités E1 et E2 sont reliées par une relation de coréférence, chaque entité hérite des relations thématiques de l’autre entité. Il n’est pas nécessaire alors d’annoter pour E1 les relations impliquant déjà E2.

Dans l’exemple ci-dessous, Fusarium oxysporum f. Sp. Cubense fungi et Foc sont reliés par une relation de coréférence. Il n’est pas nécessaire de créer une relation de cause entre le pathogène Foc et la maladie fusariosis. La relation entre Fusarium oxysporum f. Sp. Cubense fungi et Foc suffit.

Lorsque le terme coréférent n’est pas synonyme de l’antécédent, il n’est pas annoté, notamment lorsqu'il est plus général.

Par exemple bacteria n’est pas annotée en coréférence dans

Apple proliferation phytoplasma, this bacteria

|  |  |  |
| --- | --- | --- |



## Relations binaires

La relation binaire relie deux entités. Elle est orientée, elle a un sens. L’entité argument de départ est appelée source et l’entité argument d’arrivée est appelée cible. Les relations ont une étiquette associée qui caractérise le type de la relation par exemple, cause entre organisme nuisible et maladie.

Les sections ci-dessous sont organisées par type de relation binaire. Pour chacune le type des entités sources et cibles possibles sont indiquées.

### **Cause**

| Entité source | ON |
| --- | --- |
| Entité cible | Maladie |
| Définition | Relation entre un organisme nuisible et la maladie qu’il cause. |



| Exemples | | |
| --- | --- | --- |
|  |  | Foc causing fusariosis |

### **Trouve\_sur**

| Entité source | ON ou Vecteur |
| --- | --- |
| Entité cible | Plante ou Voie\_de\_dispersion |
| Définition | Relation entre un organisme nuisible ou vecteur et l’entité sur laquelle on peut le trouver, une plante ou une voie de dispersion. |

#### Exemples

|  |  | Fusarium was found on banana trees |
| --- | --- | --- |

### **Vecte\_par**

| Entité source | ON |
| --- | --- |
| Entité cible | Vecteur |
| Définition | Relation entre un organisme nuisible et un vecteur |

#### Exemples

|  |  | Candidatus Liberibacter asiaticus is vectored by Diaphorina citri Kuwayama |
| --- | --- | --- |

### **S\_exprime\_sur**

| Entité source | Maladie |
| --- | --- |
| Entité cible | Plante |
| Définition | Relation entre une plante et la maladie qui s’y exprime. |

| Exemples | | |
| --- | --- | --- |
|  |  | they help manage fusariosis in banana trees, |

### **Disperse\_par**

| Entité source | ON ou Maladie |
| --- | --- |
| Entité cible | Voie\_de\_dispersion |
| Définition | Relation entre l’organisme nuisible ou la maladie, et sa voie de dispersion. De fait les maladies ne sont dispersées en elle-mêmes, mais cet abus de langage qui identifie la maladie à l’agent est fréquent dans les textes. |



| Exemples | | |
| --- | --- | --- |
|  |  |  |
|  |  | common pathway for Oriental fruit fly to enter the State is by “hitchhiking” in fruits |

### **Localise\_en**

| Entité source | ON ou Vecteur ou Plante ou Maladie ou Voie\_de\_dispersion |
| --- | --- |
| Entité cible | Lieu |
| Définition | Relation entre une entité observée et le lieu de l’observation.  Un organisme pathogène qui provoque une épidémie est localisé dans le lieu de l'épidémie. |



| Exemples | | |
| --- | --- | --- |
|  |  | Fusariosis is common in Brazil |

### **Detecte\_le**

| Entité source | ON ou Vecteur ou Plante ou Maladie ou Voie\_de\_dispersion |
| --- | --- |
| Entité cible | Date |
| Définition | Relation entre l’entité observée et la date de l’observation. |



| Exemples | | |
| --- | --- | --- |
|  |  | HLB was reported in South China in 1943. |

### **Quantifie**

| Entité source | ON ou Vecteur ou Plante ou Voie\_de\_dispersion ou Lieu |
| --- | --- |
| Entité cible | Quantité |
| Définition | Relation entre sa quantité et une entité observée. |



| Exemples | | |
| --- | --- | --- |
|  |  | In 1948, the pine wilt disease was found on 3 pine trees in the region [..] |

## Relations n-aires

| Définition |  |  |
| --- | --- | --- |

Les relations n-aires sont des relations qui relient plus de deux arguments. Elles sont aussi appelées "événements''. Une première relation binaire relie les deux entités principales. Les autres relations relient la première relation avec les entités supplémentaires. Le sens de la relation est le même que celui des relations binaires

Le choix de la première relation est effectué selon un **ordre de priorité fixe** :

1. *Cause* [ON -> Maladie]
2. *Trouve\_sur* [ON -> Plante]
3. *Vecte\_par* ou *Disperse\_par* [ON ou Maladie -> Vecteur ou Voie\_de\_dispersion]
4. *Localise\_en* [? -> Lieu]
5. *Detecte\_le* [? -> Date]



| Exemples | | |
| --- | --- | --- |
|  |  |  |
|  | **Vection :  ON-vecteur-plante** | The banana root borer transmits Fusarium to banana trees |
|  |  |  |
|  | **Contamination : ON-maladie-plante** | Fusariosis is caused by the presence of Fusarium on banana trees |

## Modalité

La propriété modalité est associée à certaines relations. Elle peut prendre la valeur neutral par défaut, ou négation ou hypothèse.

### **Négation**

Elle prend une valeur “négation” lorsque la relation est explicitement niée dans le document.

| Exemple | | |
| --- | --- | --- |
|  | **négation** | millet, on the other hand, have not displayed potential as alternative hosts of Foc |

### **Hypothèse**

La modalité d’une relation prend la valeur “hypothèse” lorsque la relation est explicitement hypothétique dans le document.



| Exemple | | |
| --- | --- | --- |
|  | **hypothèse** | Crotalaria [..]identified as potential hosts of Foc |