# Fiche 12 - Coréférence

## Sommaire

1. Description de la tâche
   1.1. Coréférence sur INCEpTION
   1.2. Protocole
2. Retour sur expérience

**Mots-clés :** couche d'annotation, coréférence, relation

## 1. Description de la tâche

### *1.1 La coréférence sur INCEpTION :*

L'annotation de la coréférence peut se faire à l'aide d'une couche intégrée à la plateforme INCEpTION (Coreference). C'est une annotation de type Relation (cf *Fiche 7 - Les types de relations*), avec des paramètres réglés par défaut. Certains réglages sont modifiables, et d'autres non.

Paramètres pour la couche Coreference :

- *Enable* : ce paramètre offre la possibilité de supprimer la couche d'annotation. Par défaut, il est activé, mais il est possible de le désactiver, et donc d'autoriser la suppression de cette couche.

- *Read only* : ce paramètre permet de choisir si cette couche est éditable ou juste visualisable. Par défaut, il est coché, donc éditable. Cependant il est possible de le décocher pour permettre seulement la visualisation de cette couche.

- *Validation* : ce paramètre permet d'activer ou de désactiver la vérification automatique de la bonne formation des annotations (selon les autres paramètres pré-définis) au moment de la clôture du document. Par défaut, la vérification se fait à chaque fois (*Always*), mais il est possible de ne jamais la faire automatiquement (*Never*).

- *Granularity* : ce paramètre définit le niveau sur lequel l'annotation porte. Le paramètre définit est "*token-level*". Ainsi, avec ce paramètre, il est possible que plusieurs mots-formes soient intégrés dans la même couche d'annotation. Il est possible de modifier ce paramètre. En effet, il est possible de sélectionner "*Single token only*", ce paramètre imposera le choix d'un seul mot-forme par annotation. Il est également possible de choisir le paramètre "*Character-level*", ici, il permet de sélectionner un ou plusieurs caractères pour définir la coréférence dans un texte. Enfin, il est possible de choisir le paramètre "*Sentence-level*". Ce paramètre permet de sélectionner une ou plusieurs phrases pour annoter la coréférence. Il est alors vivement conseillé, pour annoter au mieux la coréférence de laisser la granularité sur "*token-level*".

- *Overlap* : ce paramètre permet ou interdit la superposition et l'enchâssement des étiquettes. Ce paramètre est réglé sur "*Any*" ce qui rend possible la superposition et l'enchaînement. Cependant, ce paramètre peut être modifiable. Il peut dans un premier temps être réglé sur "*None*", rendant impossible la superposition et l'enchassement.

- *Allow crossing sentence boundaries* : ce paramètre est par défaut coché. Il permet le franchissement des limites de la phrase. Ainsi, grâce à ce paramètre, la coréférence peut s'étendre sur plusieurs phrases.

- *Behave like a linked list* : ce paramètre est par défaut décoché. Il permet de faire en sorte que la coréférence se comporte comme un liste chaînée. Il vaut mieux ne pas le cocher.

*Pour plus de détails sur les paramètres d'une couche d'annotation, voir Fiche 8 - Création d'une couche d'annotation.*

## ***1.2 Protocole :***

Dans un discours il est possible et même certain qu'il y ait plusieurs référents, ainsi il y aura plusieurs chaînes de coréférence au sein du même discours. Pour annoter la coréférence sur INCEpTION, il est essentiel de suivre différentes étapes :

**Etape 1 :** Choisir la couche d'annotation (*layer*) "Coreference" à droite sur l'image ci-dessous

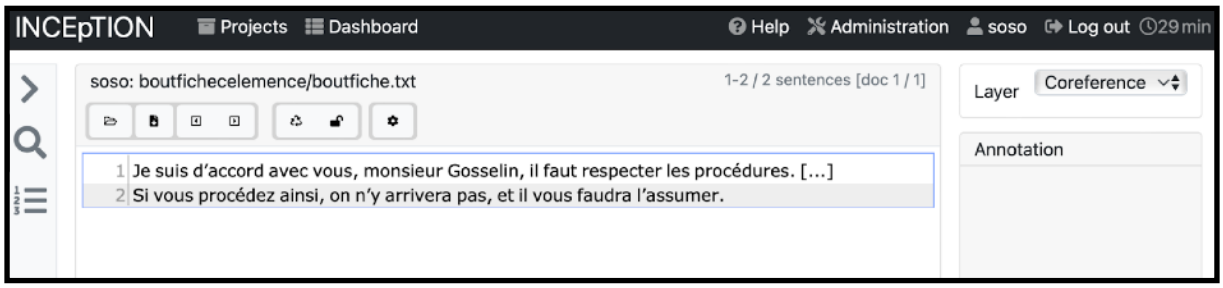


*Image 1 : segment à annoter + layer coréférence*

**Étape 2 :** Surligner la première et la deuxième expression référentielle

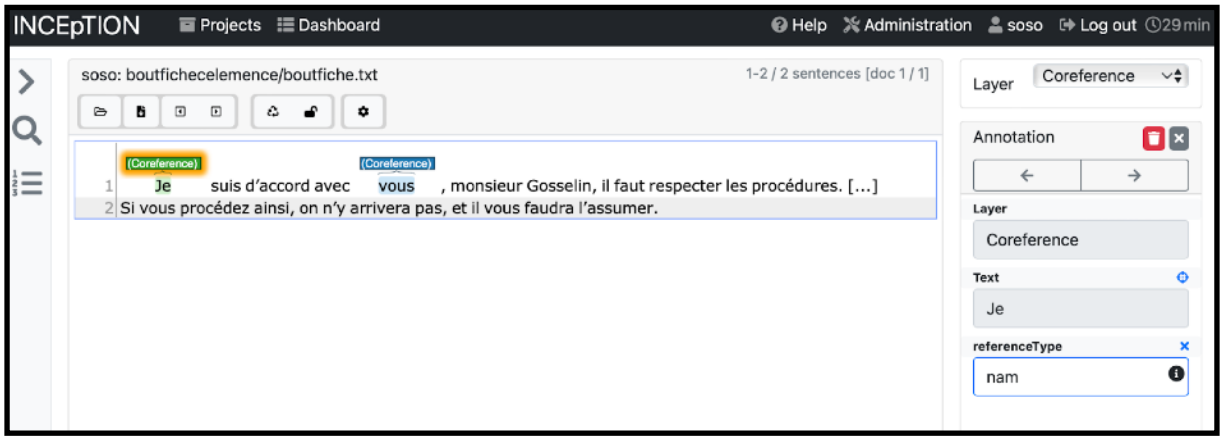


*Image 2 : sélection de deux coréférences*

**Étape 3 :** Relier les deux expressions

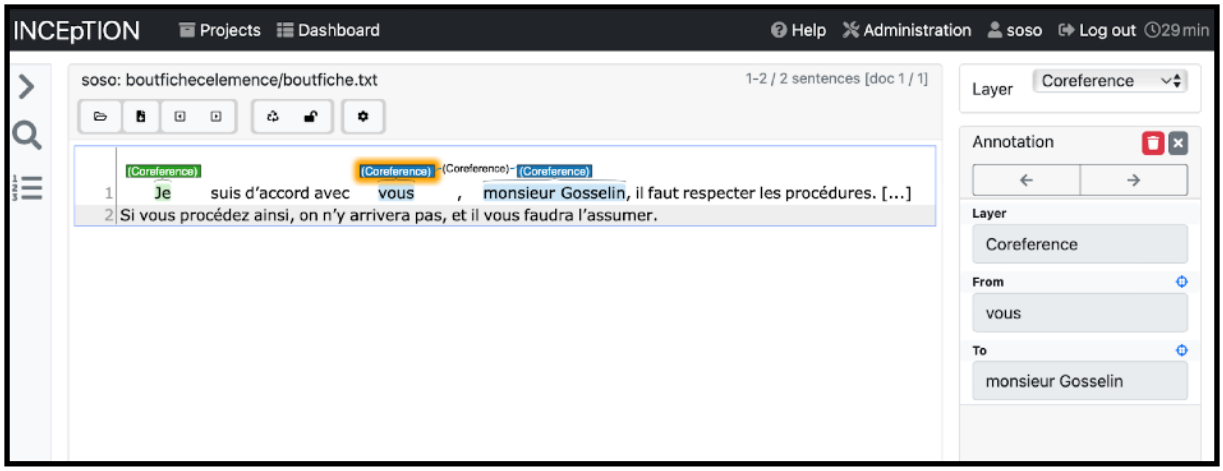


*Image 3 : une chaîne de coréférence*

Pour relier les deux expressions référentielles, il suffit de cliquer sur celle à annoter, puis d'amener la souris sur la deuxième expression référentielle.

Ces différentes étapes ont permis de créer une chaîne de référence. Il suffit de continuer sur ce même principe pour annoter toutes les chaînes de coréférence au sein du discours.

*Image 4: le segment entier annoté*

Dans l'image ci-dessus, une chaîne de référence est représentée avec la couleur bleue. Les couleurs sont utiles et permettent de bien mettre en évidence chaque chaîne de coréférence indépendamment les unes des autres.

## 2. Retour sur expérience

La tâche d'annotation concernant la coréférence n'est pas une tâche facile, puisqu'il n'y a pas forcément de bonne réponse dans l'annotation. En effet, d'un annotateur à l'autre la coréférence dans un texte peut diverger selon le point de vue. Il est alors important de prendre des décisions personnelles et de se renseigner sur ce qu'est la coréférence pour l'annoter au mieux (*cf Fiche 6 - Guide d'annotation*).

Cependant, une fois que la tâche est effectuée, INCEpTION permet de mettre en avant les différentes relations de coréférence et les différentes chaînes.

### POINTS D'AMELIORATION

Le problème qui se pose est qu'INCEpTION ne propose pas de nom pour nommer une chaîne de coréférence. Ainsi, qu'elles se référent à un objet, à une pensée, à un humain… elles portent toutes le même nom, soit coréférence.

Il se pose également un problème concernant la création de relation de coréférence quand deux éléments coréférents sont loin dans un texte. En effet, lorsque deux éléments coréférents se trouvent éloignés dans un texte, il est difficile de les annoter correctement.