

AnnoTerm

Version du 20 août 2026

Vous participez à une étude sur la façon dont les relevés terminologiques varient d'une personne à l'autre. Les annotations recueillies serviront ensuite à évaluer des systèmes d'extraction terminologique automatique.

Il n'existe pas de relevé correct pour cette tâche : nous cherchons justement à mesurer comment les décisions varient d'une personne à l'autre. Vos annotations ne seront comparées à aucun corrigé, et cet exercice n'évalue pas vos compétences.

Les annotations se feront dans la plateforme ***INCEpTION*** ; plus de détails sont disponibles plus loin dans ce document.

La Section 1 explique la tâche d'annotation, la Section 2 explique les ressources permises, la Section 3 décrit l'environnement d'annotation et répond à certaines questions récurrentes sur ***INCEpTION***. Finalement, la Section 4 mentionne quoi faire à la fin de l'annotation.

Les Sections 1, 2 et 4 sont importantes à lire avant de commencer la tâche, et nécessaires à l'annotation. La Section 3 est utile lorsque vous avez des questionnements sur la plateforme d'annotation, mais elle n'est pas nécessaire pour comprendre la tâche.

1 Tâche

Vous n'avez qu'un seul objectif pour les dix textes :

Annotez ce que vous ajouteriez à votre propre base terminologique si vous aviez à traduire d'autres textes semblables.

- Annotez **chaque occurrence** de ce qui vous semble pertinent.
 - Si une forme apparentée apparaît (un pluriel, une autre catégorie grammaticale, une variante graphique), décidez pour cette forme-là selon ses propres mérites : annotez-la si vous l'ajouteriez, laissez-la si vous ne l'ajouteriez pas. Vous n'avez pas à vous demander si deux formes constituent « le même terme » ; c'est notre problème, pas le vôtre.
- Un élément pertinent **peut ne pas être sous forme lemmatisée**.
- **Aucune limite de longueur maximale** pour un segment annoté. La plus petite unité annotable est le mot ; vous ne pouvez pas annoter une partie d'un mot.
- Aucun nombre minimal ou maximal d'éléments à annoter.
- Toutes les **catégories lexicales pleines** sont admissibles (noms, verbes, adjectifs, adverbes).

- Ceci ne concerne que le noyau syntaxique d'un segment ; un segment de plusieurs mots peut contenir des prépositions, déterminants, etc.
- Les noms propres ne font pas l'objet d'un traitement particulier ; ils peuvent être annotés si vous les jugez pertinents.
- Annotez les unités telles qu'elles apparaissent, **sans corriger les erreurs**.

Un segment doit être **continu** : il ne peut pas sauter de mots. Si l'unité qui vous intéresse est interrompue par d'autres mots (p. ex. « tenir compte » dans « Il faut tenir systématiquement compte des délais »), ne l'annotez pas, à moins que vous considériez les mots intercalés pertinents. Annotez seulement une partie continue si elle est utile en elle-même.

L'imbrication des annotations est permise sans être requise. Si un segment plus court situé à l'intérieur d'un segment déjà annoté mérite selon vous sa propre entrée, annotez-le également. Sinon, ne le faites pas.

Vous n'avez qu'à marquer les occurrences dans le texte, telles qu'elles apparaissent. Vous n'avez pas à produire la forme que vous inscririez réellement dans votre base : espaces, traits d'union, casse et autres détails de forme sont uniformisés au moment de l'analyse.

2 Ressources permises et non-discussion

Vous pouvez consulter les ressources que vous utiliseriez normalement : dictionnaires, banques de terminologie, glossaires, documentation du domaine, moteurs de recherche, traduction automatique.

Vous pouvez aussi demander à un agent conversationnel de vous expliquer une notion. Une seule exclusion : aucun outil ne doit vous fournir de candidats. Concrètement, n'utilisez pas d'extracteur de termes, ni la fonction d'extraction d'un outil de TAO, et ne demandez à aucun système de repérer, de lister ou de proposer les termes d'un texte.

La distinction est la suivante : une ressource peut vous aider à trancher au sujet d'un segment que vous avez repéré vous-même, mais elle ne doit pas repérer les segments à votre place.

Vous pouvez demander à un Grand Modèle de Langue (GML), tel que ChatGPT, Gemini ou Claude, de donner d'éclaircir certains éléments du texte (p. ex. « Quelle est la structure actantielle de “sertir” dans le domaine de la joaillerie ? », « Que signifie “lapidaire” »), mais **devez éviter** de demander à ces outils de faire l'extraction des termes à votre place (p. ex. « Quels sont les termes dans ce texte ? », « Est-ce que “lapidaire” est un terme dans ce texte ? »).

Jusqu'à ce que nous vous confirmions la fin de la collecte, n'échangez avec personne au sujet des textes, de vos décisions ou de vos hésitations : ni avec les autres personnes participantes, ni avec qui que ce soit d'autre. Une discussion, même brève, rapproche les annotations et

fausse la mesure qui nous intéresse.

À la fin de la période d’annotation, un court questionnaire vous demandera quelles ressources vous avez consultées. Répondez-y avec exactitude et honnêteté. Aucune réponse n’entraîne de conséquence pour vous ; elles nous serviront à interpréter les résultats.

3 Travailler dans *INCEpTION*

La plateforme utilisée pour l’annotation est *INCEpTION*. La Figure 1 présente l’écran de connexion.

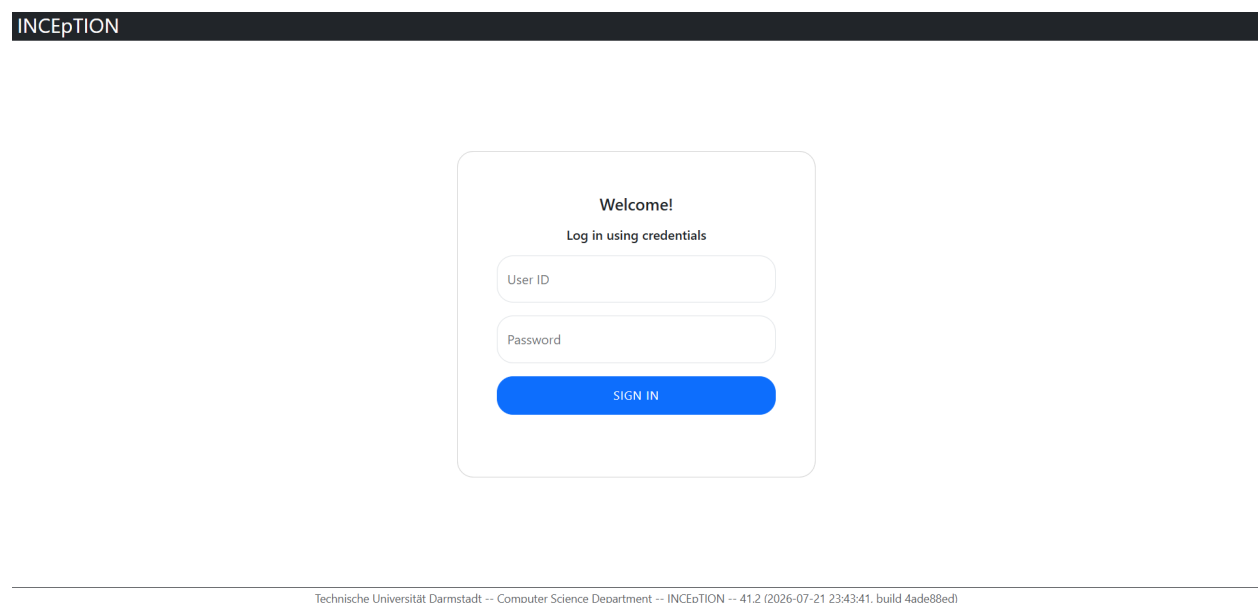


FIGURE 1 – Écran de connexion pour *INCEpTION*.

3.1 Modification du mot de passe

Lorsqu’une personne participante reçoit son invitation pour la plateforme *INCEpTION*, celle-ci doit modifier le mot de passe par défaut reçu par courriel lors de sa première connexion.

Une fois connectée, la personne annotatrice peut modifier son mot de passe en cliquant sur son nom en haut à droite de l’écran (Figure 2). L’écran suivant permettra de modifier un mot de passe. Noah peut être en mesure de réinitialiser un mot de passe au besoin.

Cliquer sur *INCEpTION* (haut gauche) permet de retourner à l’écran d’accueil.

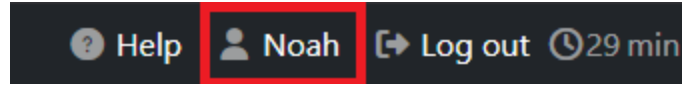


FIGURE 2 – Cliquer sur son nom (en haut à droite de l’écran).

3.2 Environnement d’annotation

À partir de l’écran d’accueil, l’option *Annotation* est accessible (à la gauche de la fenêtre). En appuyant sur cette option, le menu de sélection de documents apparaît (Figure 3). Cliquer sur un document permet d’ouvrir l’environnement d’annotation.

Choose Document ×

▼
○
⌂
⌕

State	Name ^	Tokens	Sentences	Created	Updated
☒	en_01.txt	324	15	2026-08-03	2026-08-20
☐	en_02.txt	332	9	2026-08-03	2026-08-20
☐	en_03.txt	281	11	2026-08-03	2026-08-20
☐	en_04.txt	318	12	2026-08-03	2026-08-20
☐	en_05.txt	305	11	2026-08-03	2026-08-20
☐	fr_01.txt	270	11	2026-08-03	2026-08-20
☐	fr_02.txt	338	14	2026-08-03	2026-08-20
☐	fr_03.txt	217	7	2026-08-03	2026-08-20
☐	fr_04.txt	286	10	2026-08-03	2026-08-20
☐	fr_05.txt	265	8	2026-08-03	2026-08-20

FIGURE 3 – Menu de sélection de documents.

La Figure 4 montre l’environnement d’annotation pour un document fictif. L’écran est divisé en trois sections : au centre, la zone d’annotation ; à droite, la section d’options et de détails sur un élément annoté précis ; à gauche, la section regroupant tous les éléments annotés.

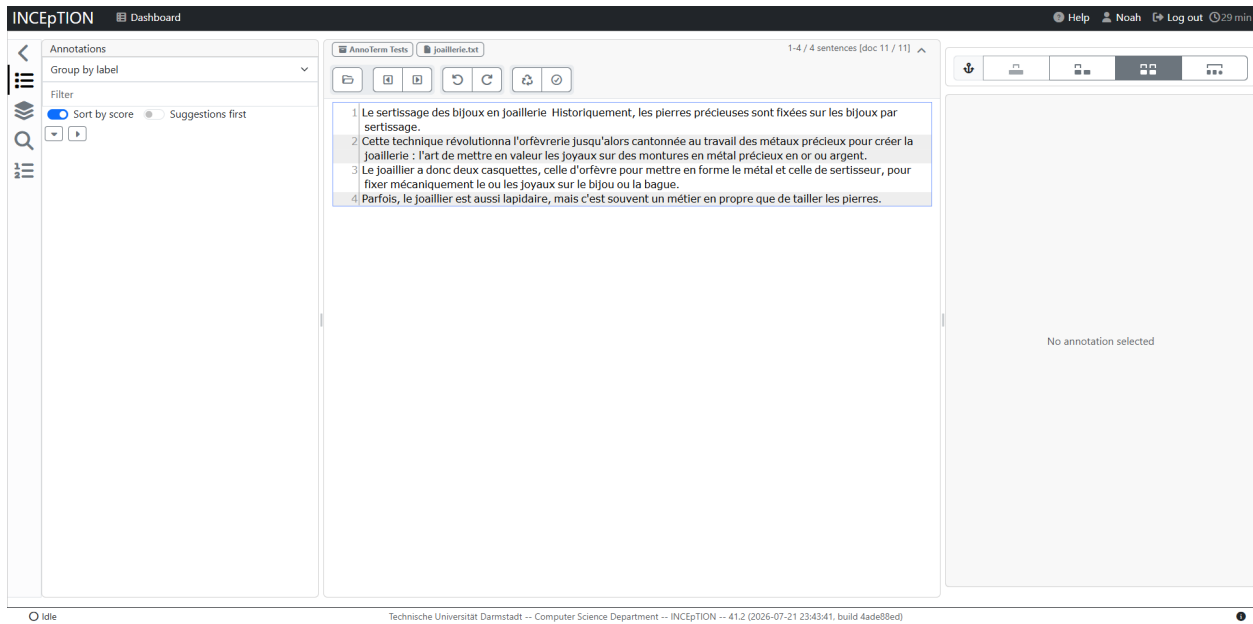


FIGURE 4 – Environnement d’annotation.

La section principale est celle au centre. Les options au-dessus de la section centrale (Figure 5) permettent, de gauche à droite :

- d’ouvrir un prochain document ;
- de naviguer entre les documents précédents et suivants (les flèches) ;
- d’annuler et de refaire une action ;
- de réinitialiser toutes les annotations du document (ceci demande ensuite une confirmation) ;
- de modifier l’état du document entre « Annotation in progress » et « Annotation finished » (l’option « Toggle state »).

Lorsque vous avez terminé l’annotation d’un document, veuillez confirmer sa complétude avec l’option « Toggle state ». C’est à vous de décider quand un document est vraiment terminé.

Vous pouvez retourner en « Annotation in progress » si vous souhaitez revisiter une annotation qui était marquée comme terminée.

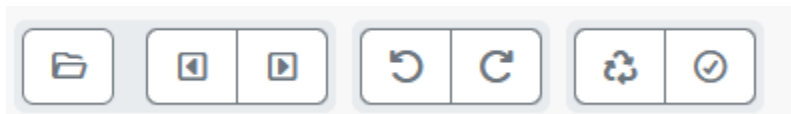


FIGURE 5 – Options au-dessus de la section centrale.

La section à droite de l’écran sert à visualiser des annotations sélectionnées, et à laisser une remarque sur un élément annoté.

La section de gauche regroupe les éléments qui furent annotés. Lorsque du texte est sélectionné dans la section centrale, ce segment est considéré comme un terme ; il apparaît alors dans la section de gauche comme le montre la Figure 6. Vous pouvez retirer un terme annoté et voir tous les termes annotés en ordre d'apparition dans le document.

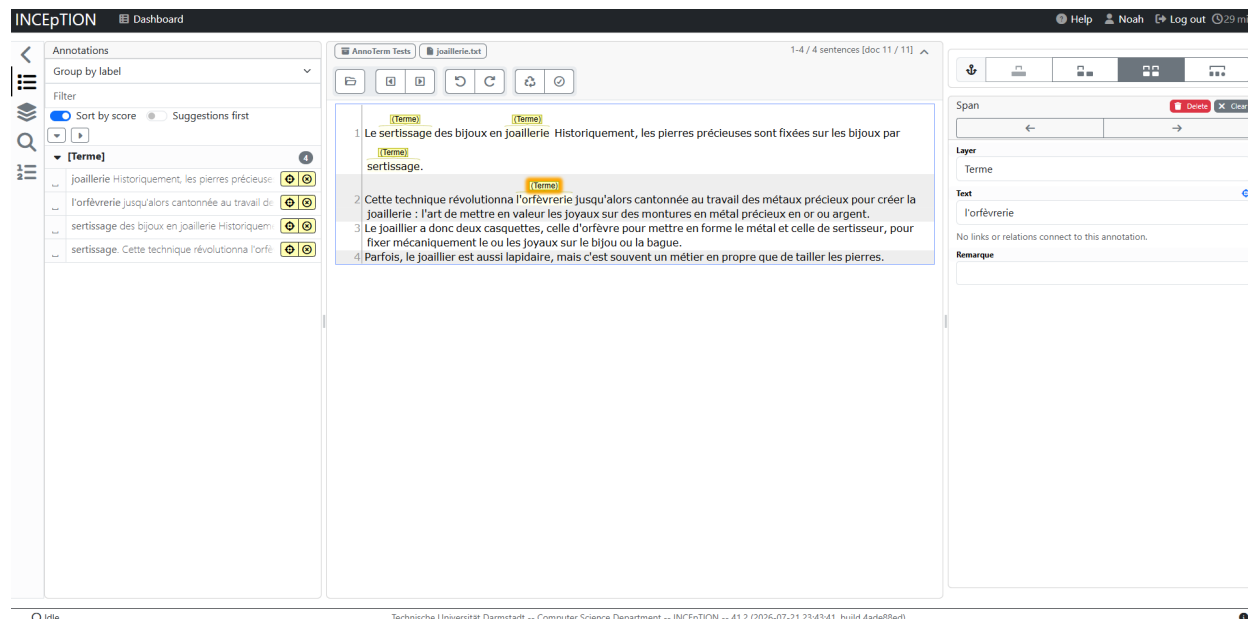


FIGURE 6 – Exemple d’annotation en cours. La section de gauche montre toutes les annotations faites en ordre d’apparition dans le texte. La section de droite montre les informations sur l’annotation sélectionnée (« l’orfèverie »), avec une option pour ajouter des remarques sur cette annotation.

Vous pouvez interrompre votre travail et reprendre plus tard ; il est enregistré automatiquement au fur et à mesure. La session expire après un certain moment d’inactivité (compte à rebours en haut de l’écran), sans perte de données ; il suffit alors de se reconnecter pour continuer l’annotation.

3.3 Frontière de mots

INCEpTION est réglé pour annoter des mots entiers ; on peut annoter un ou plusieurs mots, mais on ne peut descendre sous les frontières de mots fixés automatiquement. Par exemple, la plateforme ne permet pas d’annoter que « arc » dans la forme « arc-en-ciel », puisque les traits d’union sont considérés comme faisant partie du mot.

Certains caractères sont automatiquement rattachés au mot précédent par *INCEpTION*. Par exemple, la forme anglaise « 1.5% » est traité comme une seule unité. En revanche, la forme française « 1.5 % » serait traitée comme deux unités différentes (« 1.5 » et « % »). De plus, la plateforme considère que deux mots reliés par une apostrophe (élision ou marque du possessif) ne forment qu’une seule unité (p. ex. « l’annotation » est considéré comme un tout ; ses parties ne sont pas annotables séparément).

Dans ces cas, annotez l'unité qui vous semble pertinente telle que la plateforme la découpe. Ces détails sont uniformisés au moment de l'analyse.

3.4 Granularité

Parmi les fonctionnalités présentes dans la section de droite de l'environnement d'annotation, une option permet de modifier la granularité des annotations (Figure 7). Par défaut, la granularité est réglée sur « Token-level » (troisième option). C'est le niveau de granularité souhaité pour cet exercice. Si vous vous retrouvez à annoter au niveau du caractère ou de la phrase, veuillez resélectionner l'option « Token-level », comme dans la Figure 7.



FIGURE 7 – Options de granularité d'annotation. Ceci devrait être réglé comme dans l'image (sur l'option « Token-level »).

3.5 Ordre des documents

Vous avez dix textes à annoter ; cinq en anglais et cinq en français, provenant du domaine de l'insuffisance cardiaque. Annotez les documents dans l'ordre où ils apparaissent dans la liste de la plateforme.

4 Pour finir

Pour toute question d'ordre technique ou relative aux directives d'annotation, écrivez à Noah Tremblay Taillon (noah.tremblay.taillon@umontreal.ca).

Les questions sur les directives ne reçoivent pas de réponse individuelle. Si une clarification s'avère nécessaire, elle sera transmise à l'ensemble des personnes participantes, de sorte que personne ne dispose d'une précision que les autres n'ont pas.

Veuillez écrire à Noah lorsque vous aurez terminé l'annotation des dix documents. Vous recevrez alors le questionnaire de fin de tâche. Vous devrez notamment y insérer le temps total utilisé pour l'exercice (incluant la lecture des directives). Lorsque vous recevrez ce sondage final, vous ne serez plus en mesure de modifier vos annotations.