Formation doctorale « Explorer des données textuelles avec TXM (initiation au logiciel) »

Université de Neuchâtel, 06/11/2026

Salle 101, rue de la Pierre-à-Mazel n°7, 2000 Neuchâtel

Programme CUSO Littérature Française

Programme

9h-9h30	Accueil (café) et vérification des installations de TXM
9h30–10h30	Lire autrement : explorer un corpus avec TXM
10h30–10h45	Pause-café
10h45–12h15	Du mot aux contrastes : interroger les textes et comparer les parties du corpus
12h15–14h	Repas au Café des Amis (Quai Robert-Comtesse n°4)
14h–15h30	Des contrastes aux sens
15h30–15h45	Pause-café
15h45–17h	De l'analyse à la pratique : explorer les contextes et importer ses propres corpus

Guide d'installation de TXM sur vos machines

Nous demandons aux participant·es d'**avoir installé TXM sur leur machine avant la formation**.

La version de TXM utilisée pour la formation est la **version 0.8.4**, et non la version 0.8.5 sortie il y a peu. Nous demandons aux participant·es de bien veiller à avoir la version 0.8.4 (on peut avoir plusieurs versions de TXM sur la même machine).

En cas de difficultés avec l'installation, vous pouvez contacter Timothée Premat jusqu'au 05/11 à l'adresse timothee.premat@unine.ch. Timothée peut également vous recevoir sur rdv (bureau 110, rue de la Piere-à-Mazel n°7).

Les instructions suivantes expliquent comment installer et tester TXM sur Windows et MacOS. Si vous êtes sur Linux, la procédure devrait être analogue à celle sur MacOS : téléchargez l'archive, ouvrez-la avec un utilitaire d'installation, puis procédez comme sur MacOS.

Installer TXM 0.8.4 sur Windows

1. Installer et ouvrir TXM

Il vous faut Windows 10 ou 11 en 64 bits, une connexion Internet et les droits d'installation. Sur un ordinateur professionnel, sollicitez le service informatique si nécessaire.

Ouvrez la page de téléchargement de TXM 0.8.4 et cliquez sur **Windows 10 - 11**. Double-cliquez sur le fichier téléchargé et suivez les instructions.

Lancez TXM depuis **Démarrer** → **TXM-0.8.4** → **TXM-0.8.4**.

Attendez le message **« TXM est prêt. »** dans la zone de texte en bas de la fenêtre : c'est la **console**.

2. Ajouter la lemmatisation

Dans TXM, cliquez sur **Fichier** → **Ajouter une extension**.

Cochez **TreeTagger software** et **TreeTagger models**.

Suivez les étapes proposées, puis **redémarrez TXM**.

> Ces deux extensions installent l'outil et les modèles français et anglais. Vous n'avez pas à les télécharger séparément.

3. Faire un petit test

Sélectionnez ces phrases dans la fiche et copiez-les (Ctrl + C) :

> *Les enfants mangent des pommes. Nous mangeons ensemble.*

Dans TXM : **Fichier** → **Importer** → **Corpus** → **Presse-Papier**. Nommez le corpus **TEST**, choisissez la langue **fr** et validez.

Dans la console, cherchez **frpos** et **frlemma** dans la ligne **pAttrs**, puis le message **« Import terminé »**.

Vous voyez les deux propriétés et l'import est terminé ?
Bravo, la lemmatisation est opérationnelle !

> **Un blocage ?** Faites une capture du message et transmettez-la à l'équipe de formation avant la séance, en précisant la version de votre système d'exploitation (OS). Si frlemma manque, vérifiez les deux extensions, redémarrez TXM et refaites le test.

Sources officielles : TXM 0.8.4 · TreeTagger, section B · FAQ, vérification de l'installation

Installer TXM 0.8.4 sur votre Mac

1. Installer et ouvrir TXM

Prévoyez une connexion Internet et les droits d'installation. Le lien Mac du fournisseur mentionne les puces Intel, M1, M2 et M3. Pour une autre configuration, contactez l'équipe de formation en cas de doute.

Ouvrez la page de téléchargement de TXM 0.8.4 et cliquez sur **Mac**. Ouvrez le fichier téléchargé et suivez les instructions. Puis, dans le Finder : **Applications → TXM-0.8.4 → TXM**.

> Si macOS signale un développeur non identifié : fermez l'alerte, ouvrez les réglages **Sécurité et confidentialité** (ou **Confidentialité et sécurité**), puis cliquez sur **Ouvrir quand même** pour TXM.
> Le message d'avertissement sur la prise en charge des puces Intel par MacOS peut être ignoré.

Attendez le message « TXM est prêt. » dans la zone de texte en bas de la fenêtre : c'est la console.

2. Ajouter la lemmatisation

Dans TXM, cliquez sur **Fichier → Ajouter une extension.**

Cochez **TreeTagger software** et **TreeTagger models**.

Suivez les étapes proposées, puis **redémarrez TXM**.

> Ces deux extensions installent l'outil et les modèles français et anglais. Vous n'avez pas à les télécharger séparément.

3. Faire un petit test

Sélectionnez ces phrases dans la fiche et copiez-les (Ctrl + C) :

> *Les enfants mangent des pommes. Nous mangeons ensemble.*

Dans TXM : **Fichier → Importer → Corpus → Presse-Papier**. Nommez le corpus **TEST**, choisissez la langue **fr** et validez.

Dans la console, cherchez **frpos** et **frlemma** dans la ligne **pAttrs**, puis le message **« Import terminé »**.

Vous voyez les deux propriétés et l'import est terminé ?
Bravo, la lemmatisation est opérationnelle !

> **Un blocage ?** Faites une capture du message et transmettez-la à l'équipe de formation avant la séance, en précisant la version de votre système d'exploitation (OS). Si frlemma manque, vérifiez les deux extensions, redémarrez TXM et refaites le test.
>
> Sources officielles : TXM 0.8.4 · TreeTagger, section B · FAQ, vérification de l'installation

Organisateur·trices

Nina Mueggler	UniNe, Professeure assistante	IFL, chaire de litt. fr. XVI-XVII^{e} s.	nina.mueggler@unine.ch
Timothée Premat	UniNe, SNSF SPF Fellow	ISLa, chaire de dialectologie	timothee.premat@unine.ch
Céline Rumpf	UniNe, Assistante doctorante	ISLa, chaire de dialectologie	celine.rumpf@unine.ch

Formateur·trices

Hugo Dumoulin	Université Paris Nanterre & CNRS	MoDyCo	hdumoulin@parisnanterre.fr
Virginie Lethier	Université Marie et Louis Pasteur	Elliadd	virginie.lethier@univ-fcomte.fr
Cyrielle Montrichard	Université Paris-Est Créteil	Céditec	cyrielle.montrichard@u-pec.fr

Description

Le logiciel TXM est une plateforme gratuite et open source de référence pour l'exploration de corpus textuels par les études littéraires et les disciplines de SHS (littérature, linguistique, sociologie, histoire, information-communication, géographie, sciences politiques, etc.). La formation sera dédiée à la constitution, l'import et l'exploration de corpus littéraires.

Cette formation tournée vers la découverte des fonctionnalités principales de TXM est accessible sans connaissances préalables. Une fois les corpus importés et annotés dans TXM, les participant·es seront introduit·es aux fonctions de requête (par exemple le classement des mots par fréquence), à la logique du retour au texte, aux cooccurrences (association de mots) et aux approches contrastives, qui visent à faire émerger des différences entre deux corpus ou parties d'un même corpus.

Ces méthodes sont essentiellement transférables : elles s'appliquent aussi bien à l'analyse littéraire qu'à l'analyse linguistique où à l'utilisation de sources textuelles pour la recherche en SHS. Tout·e chercheur·euse utilisant un ou plusieurs corpus délimité(s) trouvera des outils lui permettant de rechercher efficacement des configurations linguistiques dans son corpus, et des outils lui permettant de faire contraster différentes parties de son corpus (différentes périodes, différents auteurs, différents acteurs, différents sujets, etc.). À l'issue de la formation, et avec l'aide de la documentation qui leur sera fournie, les participant·es seront en mesure d'importer leur propre corpus (selon la disponibilité des sources numériques) et de l'explorer par eux-mêmes pour enrichir leur recherche.