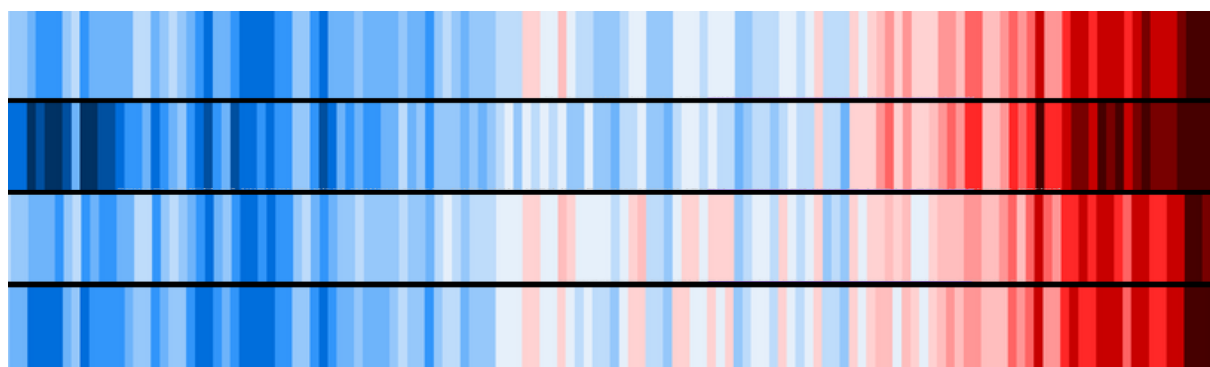


[Français ci-dessous]

## Workshop: Introduction to automatic collation



This **two-day** course offers an **introduction to the theory and the practice of automatic collation**.

**Collation** is the practice of **comparing texts** in order to identify their variant forms. It is used in the preparation of critical editions, as well in genetic criticism and in the study of textual transmissions. More in general, the **alignment** of two or more texts, which is the basis of collation, can be useful for analysing text reuse and intertextuality. Computers have been used to automatize collation since the 60's, but even though several examples of stable software are available nowadays, this tedious task is still often done by hand. This workshop will show participants how to take advantage of automatic collation tools in the field of textual and genetic criticism. A brief **history** of collation together with a reflection about the **practices** of the research community in Humanities and Social Sciences when working on the digital ecosystem will contextualize the analysis of the tools.

During the course, we will address the basic principles of the theory of automatic collation and briefly present a number of programs. In the hands-on part of the course, we will use CollateX, a collation software, to put into practice what the participants have learned. Eventually, the last part of the course can be devoted to find the right software and parameters for the use cases brought by the participants.

## Programme

### Day 1

#### **9h30-9h45 Reception**

**9h45-11h00 First session.** Explore your computer (command line, Jupyter Notebooks, Python)

#### **11h00 – 11h30 Coffee break**

**11h30 – 12h30 Second session.** [Continuation] Explore your computer (command line, Jupyter Notebooks, Python)

#### **12h30 – 14h Lunch**

**14h – 15h15 Third session.** Theory of automatic collation and reflection about the use of digital tools for textual criticism.

#### **15h15 – 15h45 Coffee break**

**15h45 – 17h00 Fourth session.** [Continuation] Theory of automatic collation and reflection about the use of digital tools for textual criticism.

**17h15 Keynote.** Elli Bleeker, Royal Academy of Arts and Sciences (The Netherlands).

**Evening.** Social dinner.

## Jour 2

**9h30 – 10h45 First session.** Automatic collation practise: environment configuration, making a simple collation.

**10h45 – 11h15 Coffee break**

**11h15 – 12h30 Invited speaker:** Rudolf Mahrer (UNIL) will present the platform *Variance*. **Second session.** [Continuation] Automatic collation practise: making a simple collation.

**12h30 – 14h00 Lunch**

**14h – 15h15 Third session.** Automatic collation practise: making a simple collation with the possibility to work with the texts of the participants.

**15h15 – 15h45 Coffee break**

**15h45 – 17h00 Fourth session.** Automatic collation practise: making a simple collation with the possibility to work with the texts of the participants.

## For whom

**The course is open to students and researchers working with texts in any discipline.**

**Requirements.** No prior knowledge of automatic collation nor programming is required.

**Your materials.** Participants will have the opportunity to send their texts in advance, so they can benefit from a personalized support during the course.

**Installation.** There is no need to install software to follow the course, because we will be using virtual machines. If, nevertheless, participants want to have their own installation, they may follow the instructions that they will receive after registration.

## Registration

Place: University of Lausanne

Date: 26-27 March 2019

PhD students of the universities of Berne, Fribourg, Genève, Neuchâtel and Lausanne (UNIL et EPFL) have priority, but postdocs and master students from Swiss universities and abroad are welcome. The participation is free, but registration is requested. For PhD students from partner universities, transportation expenses will be reimbursed together with accommodation rates for one night for participants who live more than one hour away from Lausanne. A social dinner is planned at the end of the first day.

Language of the course: English and French.

Registration: [Alexandre.Camus@unil.ch](mailto:Alexandre.Camus@unil.ch)

## Speakers

**Helena Bermúdez Sabel** is a post-doctoral researcher in computational and historical linguistics at the University of Lausanne. She has taught different subjects related to digital philology in the following degrees:

- Master in Digital Humanities, Universidad Nacional de Educación a Distancia (UNED), 2018-2019.
- Professional Academic Expert in Digital Humanities, UNED, 2016-2018.
- Professional Academic Expert in Digital Edition, UNED, 2016-2017.

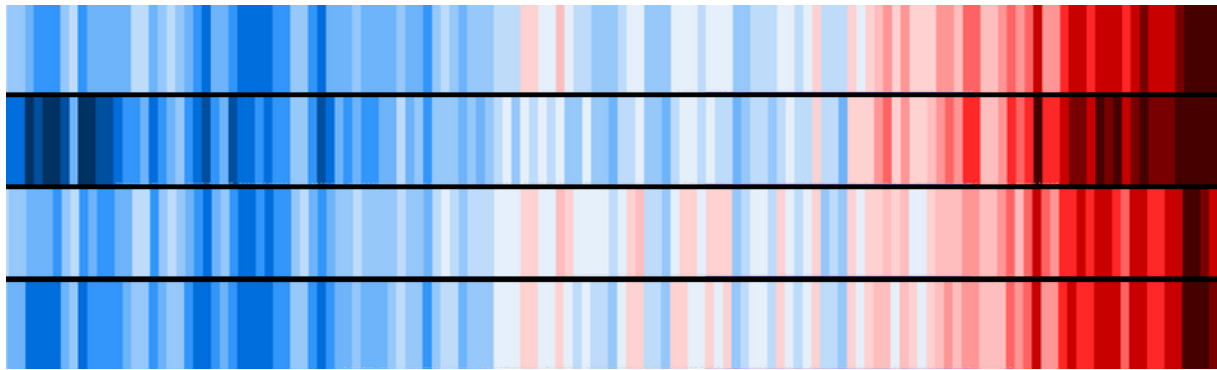
**Elisa Nury** is a post-doctoral researcher in digital papyrology at the University of Geneva. Her background is in digital philology. She has taught automatic collation in the following courses:

- École d'Été Éditions Numériques / Humanités Numériques (EDEEN), Grenoble, 28 mai - 02 juin 2018. <<https://eedeen.sciencesconf.org/>>
- Medieval and Modern Manuscript Studies in the Digital Age (MMSDA), Cambridge/London, 2-6 mai 2016. <<http://dixit.uni-koeln.de/mmsda/>>

**Elena Spadini** is a post-doctoral researcher in digital philology at the University of Lausanne. She has taught automatic collation in the following courses:

- Textes et éditions numériques, Programme doctoral en études numériques. Université de Lausanne, 25-26 avril 2019. <[https://github.com/gabays/UniL\\_DigitalTexts\\_2019](https://github.com/gabays/UniL_DigitalTexts_2019)>
- DiXiT Workshop Code and Collation: Training Textual Scholars. Amsterdam, November 2016. <<https://sites.google.com/site/dixitcodingcollation/>>

# Atelier : Introduction à la collation automatique



Ce cours de 2 jours offre une **introduction à la théorie et à la pratique** de la **collation automatique**.

La **collation** est la pratique qui consiste à **comparer** des textes afin d'identifier leurs variantes. Il s'agit d'une étape importante lors de la préparation d'une édition critique ou bien d'un dossier génétique, et permet l'étude de la transmission textuelle. Plus généralement, l'**alignement** de plusieurs textes en parallèle peut également s'avérer utile pour détecter des instances de réutilisation du texte, ou pour étudier l'intertextualité. Depuis plus de 50 ans, les chercheurs se sont efforcés d'**automatiser** la collation : aujourd'hui il existe de très bons **outils** numériques, mais cette tâche fastidieuse reste souvent effectuée à la main. Cet atelier permettra aux participants de tirer le meilleur parti des outils de collation automatique. Une brève **histoire** de la collation automatique et une réflexion sur les **pratiques** des chercheurs et chercheuses en humanités et sciences sociales qui se confrontent avec le numérique permettront de contextualiser l'étude des outils.

Pendant le cours, nous aborderons les principes de base de la théorie de la collation automatique et présenterons brièvement un certain nombre de logiciels. Dans la partie pratique du cours, nous utiliserons CollateX pour mettre en pratique ce que les participants ont appris. Finalement, la dernière partie du cours pourra être consacrée à la recherche du logiciel et des paramètres appropriés pour les cas d'utilisation apportés par les participants.

## Programme

### Jour 1

**9h30-9h45** Accueil.

**9h45-11h00 Session 1.** Explore ton ordinateur (command line, Jupyter Notebooks, Python).

**11h00 – 11h30** Pause-café.

**11h30 – 12h30 Session 2.** Explore ton ordinateur (command line, Jupyter Notebooks, Python).

**12h30 – 14h** Repas.

**14h – 15h15 Session 3.** Théorie de la collation automatique et réflexion sur l'utilisation des outils numériques dans le cadre de la critique textuelle

**15h15 – 15h45** Pause-café.

**15h45 – 17h00 Session 4.** Théorie de la collation automatique et réflexion sur l'utilisation des outils numériques dans le cadre de la critique textuelle.

**17h15 Keynote.** Elli Bleeker (Royal Netherlands Academy of Arts and Sciences) : *What has been left out: The perks and pitfalls of XML-aware collation.*

**Soirée.** Repas.

## Jour 2

**9h30 – 10h45 Session 1.** Pratique de la collation automatique : configuration de l'environnement, effectuer une collation simple.

**10h45 – 11h15** Pause-café.

**11h15 – 12h30 Session 2.** Intervention de **Rudolf Mahrer** (UNIL) sur la plateforme *Variance*.

Pratique de la collation automatique : configuration de l'environnement, effectuer une collation simple.

**12h30 – 14h00** Repas.

**14h – 15h15 Session 3.** Pratique de la collation automatique : effectuer une collation complexe avec la possibilité de travailler sur les textes des participants.

**15h15 – 15h45** Pause-café.

**15h45 – 17h00 Session 4.** Pratique de la collation automatique : effectuer une collation complexe avec la possibilité de travailler sur les textes des participants.

## Public visé

Le cours est ouvert aux étudiants et chercheurs de toute discipline qui travaillent avec des textes.

Aucune connaissance de la collation automatique ou de programmation n'est requise.

**Matériel.** Les participants ont la possibilité d'envoyer leurs textes en avance, afin de bénéficier d'un accompagnement personnalisé tout au long du cours.

**Installation.** Il n'est pas nécessaire d'installer un logiciel pour suivre ce cours, car nous utiliserons une machine virtuelle. Pour ceux qui souhaiteraient installer le programme sur leur propre ordinateur, nous enverrons des instructions aux participants après leur inscription.

## Inscription

Lieu : Université de Lausanne

Date : 26-27 Mars 2019

Les doctorants des universités de Bâle, Berne, Fribourg, Genève, Neuchâtel, et Lausanne (UNIL et EPFL) sont prioritaires, mais les postdoctorants et les étudiants (motivés) de master, en Suisse ou à l'étranger, sont les bienvenus. La participation est gratuite sur inscription. Pour les doctorants des universités partenaires, les frais de transports sont remboursés, ainsi que la nuit d'hôtel pour les participants domiciliés à plus d'une heure de Lausanne. Un repas est prévu le premier soir.

Langue d'enseignement : anglais and français.

Inscription : [Alexandre.Camus@unil.ch](mailto:Alexandre.Camus@unil.ch)

## Keynote

Elli Bleeker works as postdoctoral researcher in the Research and Development Team at the Humanities Cluster, part of the Royal Netherlands Academy of Arts and Sciences. She specializes in digital scholarly editing and computational philology, with a focus on modern manuscripts, genetic criticism and semi-automated collation. As a Research Fellow in the Marie Skłodowska-Curie funded network DiXiT (2013 – 2017), she received advanced training in manuscript studies, text modeling, and XML technologies. She also enjoys spending time in archives looking for the forgotten writers from the 20th century.

## Enseignantes

**Helena Bermudez Sabel** est chercheuse post-doctorante en linguistique computationnelle à l'Université de Lausanne. Elle a enseigné différents aspects de la philologie numérique dans les programmes suivants :

- Master in Digital Humanities, Universidad Nacional de Educación a Distancia (UNED), 2018-2019.
- Professional Academic Expert in Digital Humanities, UNED, 2016-2018.
- Professional Academic Expert in Digital Edition, UNED, 2016-2017.

**Elisa Nury** est chercheuse post-doctorante en à l'Université de Genève. Sa thèse porte sur la collation automatique et les éditions critiques numérique. Elle a enseigné la collation automatique dans les cours suivants :

- École d'Été Éditions numériques / Humanités numériques (EDEEN), Grenoble, 28 mai - 02 juin 2018. <<https://edeen.sciencesconf.org/>>
- Études de manuscrits médiévaux et modernes à l'ère numérique (MMSDA), Cambridge / Londres, 2-6 mai 2016. <<http://dixit.uni-koeln.de/mmsda/>>

**Elena Spadini** est chercheuse post-doctorante en philologie numérique à l'Université de Lausanne. Elle a enseigné la collation automatique dans les cours suivants :

- Textes et éditions numériques. Programme de doctorat en études numériques. Université de Lausanne, 25-26 avril 2019. <[https://github.com/gabays/UniL\\_DigitalTexts\\_2019](https://github.com/gabays/UniL_DigitalTexts_2019)>
- DiXiT Workshop Code and Collation: Training Textual Scholars. Amsterdam, novembre 2016. <<https://sites.google.com/site/dixitcodingcollation/>>