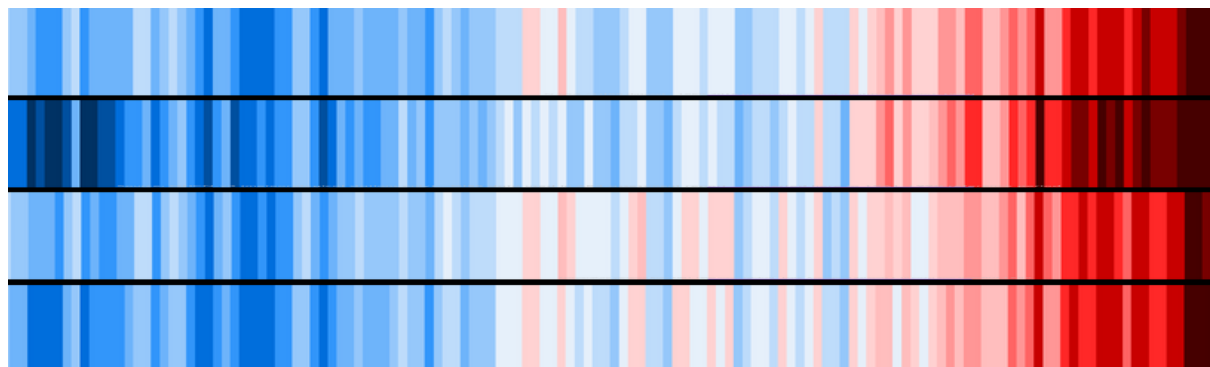


Online Workshop: Introduction to automatic collation



This **two-day** online 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 ecosystems will contextualise the analysis of the tools.

During the course, we will address the 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 Break

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

12h30 – 14h Break

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

15h15 – 15h45 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): *What has been left out: The perks and pitfalls of XML-aware collation.*

Jour 2

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

10h45 – 11h15 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 Break

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

15h15 – 15h45 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 researchers working with texts in any discipline.

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

Logistics. The course will be held **live** on an **online** platform of the University of Lausanne. Participants will receive instructions about where and when to connect. A chat and a forum will be available. A stable internet connection is needed to follow the course.

Your texts. Participants will have the opportunity to send their texts in advance, in order to benefit from a personalised support during the course.

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

Course materials. The syllabus will be available to the participants immediately before the course.

Registration

Place: online.

Date: 24-25 September 2019.

PhD students of the universities of Berne, Fribourg, Genève, Neuchâtel and Lausanne (UNIL and EPFL) have priority, but researchers of any level and master students from Swiss universities and abroad are welcome. The participation is **free**, but **registration** is requested.

Languages of the course: English and French.

Registration: 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.

Invited speaker

Rudolf Mahrer teaches French linguistics and textual genetics at the University of Lausanne. He is also editor-in-chief of *Genesis*, a reference journal in the field of genetics of works. Together with Joël Zufferey, he manages the Variance collection.

Instructors

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.

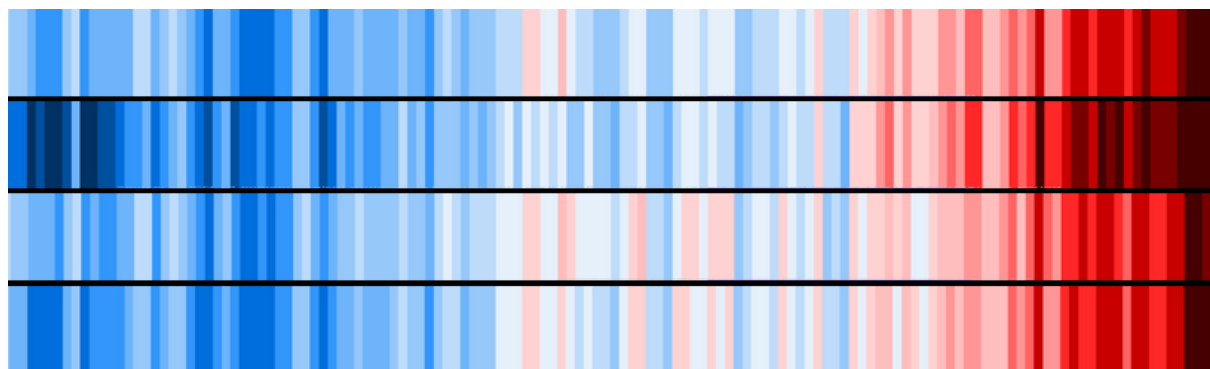
Elisa Nury is a post-doctoral researcher in digital papyrology at the University of Geneva. Her background is in digital philology. She has taught automatic collation in the following courses:

- École d'Été Éditions Numériques / Humanités Numériques (EDEEN), Grenoble, 28 mai - 02 juin 2018. <<https://edeen.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>
- DiXiT Workshop Code and Collation: Training Textual Scholars. Amsterdam, November 2016. <<https://sites.google.com/site/dixitcodingcollation/>>

Atelier en ligne : Introduction à la collation automatique



Ce cours en ligne 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 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.

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

12h30 – 14h Pause.

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.

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.*

Jour 2

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

10h45 – 11h15 Pause.

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 Pause.

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.

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 chercheurs et chercheuses de toute discipline qui travaillent avec des textes.

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

Logistique. Le cours sera donné en direct en ligne, en utilisant une des plateformes mises à disposition par l'Université de Lausanne. Les participants recevront toutes les instructions pour savoir quand et comment se connecter. Un chat et un forum de discussion seront utilisés aussi. Une connexion internet stable est requise pour suivre le cours.

Vos textes. 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 et nous expliquerons comment le faire pendant le cours. Pour ceux qui souhaiteraient installer le programme sur leur propre ordinateur, nous enverrons des instructions aux participants après leur inscription.

Matériaux de cours. Le syllabus sera mis à disposition des participants au début du cours.

Inscription

Lieu : en ligne.

Date : 24-25 septembre 2019.

Les doctorants des universités de Bâle, Berne, Fribourg, Genève, Neuchâtel, et Lausanne (UNIL et EPFL) sont prioritaires, mais les chercheurs et chercheuses de tout niveau, ainsi que les étudiant·e·s de master, en Suisse ou à l'étranger, sont les bienvenus. La participation est **gratuite, mais sur inscription**.

Langues d'enseignement : anglais et français.

Inscription : Alexandre.Camus@unil.ch

Keynote

Elli Bleeker travaille comme chercheuse post-doctorante dans l'équipe Recherche Et Développement du Humanities Cluster, à l'Académie royale néerlandaise des arts et des sciences (KNAW). Elle est spécialisée dans l'édition critique numérique et la philologie numérique, et s'intéresse en particulier aux manuscrits modernes, à la critique génétique et à la collation automatique. En participant au programme DiXiT (2013 – 2017) financé par les bourses Marie Skłodowska-Curie, elle a reçu une formation avancée sur l'étude des manuscrits, la modélisation textuelle et les technologies XML. Elle aime également passer du temps dans les archives à la recherche des écrivains oubliés du XXe siècle.

Conférencier invité

Rudolf Mahrer enseigne la linguistique française et la génétique textuelle à l'Université de Lausanne. Il dirige la revue *Genesis* et gère avec Joël Zufferey la collection *Variance*.

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>
- DiXiT Workshop Code and Collation: Training Textual Scholars. Amsterdam, novembre 2016. <<https://sites.google.com/site/dixitcodingcollation/>>