

Méthodes d'approche, de traitement et d'exploitation des grands corpus manuscrits en littérature arabe

Responsable

Bushra Farah

(Doctorante, Sorbonne
Université / Département d'Études
Arabes, Lyon 2)

Mercredi 12 juillet 2023
8h30-10h30 / 11h-13h
Salle Clio 003

Discutant

Éric Vallet

(Professeur en études arabes,
Université de Strasbourg)

Intervenants

Ibrahim Akel

(Université Sorbonne Nouvelle /
Paris 3)

Youssef Al Hamada

(Professeur agrégé, Lycée
Louis-Le-Grand, chaire supérieure
d'arabe)

Géraldine Chatelard

(Chargée de mission pour les relations
avec le Moyen-Orient, cheffe de projet
« Bibliothèques d'Orient », BnF)

Khalid Chakor-Alami

(Post-doctorant, BnF)

Bushra Farah

(Doctorante, Sorbonne
Université / Département d'Études
Arabes, Lyon 2)

Zuzana Gasakova

(Comenius University Bratislava)

Iyas Hassan

(Professeur de littérature arabe
médiévale, Sorbonne Université /
Institut français du Proche-Orient)

Résumé de l'atelier

Les deux dernières décennies ont connu une grande expansion de la recherche sur la littérature populaire arabe, notamment sur les romans épiques (*sīras*). Ceci allait de pair avec un développement très rapide de la numérisation et la diffusion des manuscrits arabes dans les différentes bibliothèques, musées et institutions académiques du monde entier. Les chercheurs et chercheuses se retrouvent de plus en plus souvent face à des corpus d'ampleur conséquente qui les interpellent quant à l'adéquation de leurs méthodes d'approche et d'analyses de leur objet. En effet, alors que l'obstacle que représentait au ^{xx}e siècle la difficulté d'accéder au manuscrit tombe, ne serait-ce que partiellement, l'abondance des sources manuscrites disponibles en ligne crée aujourd'hui de nouvelles difficultés d'ordre à la fois méthodologie et épistémologique. Que faire de cette matière colossale désormais accessible à quiconque voudra s'en servir ? Pour une même *sīra*, on peut facilement avoir à portée de main de très nombreuses copies. Pour les *sīras* les plus longues, chaque copie se présente sous forme d'une série de manuscrits, homogènes ou hétérogènes, pouvant représenter plusieurs dizaines de milliers de folios. Qu'elles soient analytiques ou comparatives, codicologiques ou historiques, narratives ou philologiques, les études réalisées sur ces corpus, surtout celles qui nécessitent l'accès à la totalité du texte, exigent des méthodes solides et efficaces pour manipuler ces nouveaux objets de recherche.

Le présent atelier réunit des praticiennes et praticiens ayant traité des manuscrits aussi bien dans l'avant et l'après la transition numérique. Il se présente ainsi à la fois comme un retour d'expérience et comme un espace de réflexion. Il abordera l'impact des dernières avancées techniques sur la recherche en littérature arabe et sur les habitudes de travail des spécialistes. Il réservera en même temps une place particulière aux outils méthodologiques à développer, voire à inventer, dans les années à venir pour aborder, traiter et exploiter les grands corpus manuscrits, tout en tenant compte de l'impact que cela pourra avoir sur la manière de penser la formation à la recherche dans ce domaine.

Pour voir l'autre face de la problématique, participent à cette discussion des bibliothèques qui ont mené ou qui mènent des opérations de numérisation d'envergure, afin de questionner les stratégies en place, leurs objectifs, leurs résultats et surtout leurs liens éventuels avec le monde de la recherche.

Programme

Ibrahim Akel

Les manuscrits de la littérature médiane aujourd'hui conservés en Syrie
L'objectif de cette intervention est de tracer, dans un premier temps, un état des lieux rapide de la littérature arabe médiane, à travers les principales collections syriennes (leur création, leur circulation et leur dispersion). Dans un deuxième temps, nous présenterons les résultats de notre recherche sur les manuscrits qui ont échappé au dépeçage des collectionneurs occidentaux ; pour fournir, enfin, une vue d'ensemble des

manuscrits de la littérature médiane encore conservés dans les bibliothèques syriennes (localisation, catalogage, accessibilité, conditions de reproduction et types d'utilisation).

Median Literature Manuscripts Preserved in Syria Today

Firstly, the goal of this intervention is to track, a rapid inventory of median Arabic literature, through the main Syrian collections (their creation, circulation and dispersion). Secondly, we will present the results of our research on manuscripts that have escaped the dismemberment of Western collectors. Those very results of our research will finally provide an overview of the manuscripts of median literature which are still preserved in Syrian libraries (location, cataloguing, accessibility, conditions of reproduction and types of use).

Youssef Al Hamada

Sīrat al-amīra Ḍāt al-himma. Les premiers pas sur le long chemin d'une édition savante

Sīrat al-amīra Ḍāt al-himma, connue également sous le titre *Sīrat al-mujāhidīn wa abṭāl al-muwaḥidīn*, fait partie de ce que l'on a coutume d'appeler la littérature populaire arabe. À l'instar de plusieurs *sīras*, celle de *Ḍāt al-himma* souffre de l'absence d'une édition savante. Dans l'objectif de parvenir à réaliser une telle édition, nous avons commencé un premier travail de tri, de classification et de comparaison à partir des différents volumes dont disposent le département des manuscrits de la BnF. Notre intervention a pour but de dresser un premier bilan de nos recherches. Nous y mettons l'accent sur quelques éléments méthodologiques auxquels nous avons eu recours lors du traitement des différents manuscrits sur lesquels nous avons travaillé.

Sīrat al-amīra Ḍāt al-himma. The first steps on the long road to scholarly publishing

Sīrat al-amīra Ḍāt al-himma, also known as *Sīrat al-mujāhidīn wa abṭāl al-muwaḥidīn*, is part of what is commonly called popular Arabic literature. Unfortunately, several *siras* have never been available in a critical edition with notes and commentary. And that includes *Sīrat al-amīra Ḍāt al-himma*.

To fill that gap, we have laid the groundwork for a complete critical edition. We have been sorting, classifying and comparing the primary sources which are housed in the BnF's department of manuscripts. Our intervention aims to draw up an initial assessment of our research. We emphasize some methodological elements that we have used during the treatment of the various manuscripts on which we have worked.

Géraldine Chatelard et Khalid Chakor-Alami

Politiques de numérisation et de diffusion numérique de la BnF et recherche sur les manuscrits arabes : défis et perspectives de collaboration

Cette communication part de deux constats nés d'échanges récents entre l'équipe de « Bibliothèques d'Orient » – collection thématique de Gallica, la bibliothèque numérique de la BnF et de ses partenaires, portant sur le Moyen-Orient et en voie de s'ouvrir au Maghreb –, et des chercheur-euses qui utilisent régulièrement les sources manuscrites numérisées par la BnF.

D'une part, à ce jour, les corpus (manuscrits ou imprimés) de littérature arabe ne font que très marginalement l'objet d'un regroupement thématique ou chronologique au sein de « Bibliothèques d'Orient », et ce alors même que l'outil de recherche de Gallica ne permet que difficilement le repérage de ces corpus par les chercheur-euses. D'autre part, la politique de numérisation de la BnF ne considère pas ces corpus comme faisant l'objet d'un traitement prioritaire alors que le milieu de la recherche sur le monde arabe y accorde un intérêt croissant. En lien avec ce dernier point, la BnF n'a pas intégré jusqu'ici dans sa chaîne de numérisation l'HTR pour les graphies non-latines.

Se pose donc la question de la place des collections arabes en général, et des manuscrits en particulier, dans les politiques de numérisation de la BnF en tant que grande bibliothèque patrimoniale à vocation universelle, ainsi que des passerelles entre le monde de la bibliothéconomie et celui de la recherche.

Le projet « Bibliothèques d'Orient », qui se caractérise par la mise en valeur thématique des collections de la BnF et de partenaires – bibliothèques patrimoniales et de recherche, fonds d'archives – situés en France et au Moyen-Orient, pourrait être un levier pour améliorer les échanges entre conservateur-trices et chercheur-euses et mobiliser des ressources (financières, scientifiques, technologiques) nécessaires à une meilleure visibilité et accessibilité numérique des corpus de littérature arabe. Cette présentation souhaiterait lancer une discussion sur le type de dispositif qui pourrait être mis en place pour améliorer la communication entre deux milieux complémentaires et mutualiser les ressources afin de mieux répondre aux attentes de la recherche dans le domaine spécifique de l'accès numérique aux manuscrits arabes tout en prenant en compte les contraintes opérationnelles de la BnF et de ses partenaires.

Digitisation and Digital Dissemination Policies at the National Library of France (BnF), and Research on Arabic Manuscripts: Challenges and Prospects for Collaboration

This presentation draws on observations made during recent exchanges between, on the one hand, the team of "Libraries of the Middle East" – a thematic collection of Gallica, the digital library of the BnF and its partners, focusing on the Middle East and in the process of opening up to the Maghreb – and, on the other hand, researchers who regularly use the manuscripts digitised by the BnF as sources.

Firstly, only a marginal number of printed or manuscript Arabic literature collections are grouped thematically or chronologically in "Libraries of the Middle East", whereas the search tool available on Gallica does not allow to locate them easily as corpuses. Secondly, in its digitisation policy, the BnF does not prioritise its Arabic literature holdings, even though Arabic studies scholars express a growing interest for this field of inquiry. Furthermore, HTR for non-Latin scripts is not yet integrated into the digitisation process of the BnF. This raises the question of the place of Arabic collections, and manuscripts, in particular, in the digitisation policies of the BnF as a major heritage library with a universal mission, as well as of the bridges between the domains of librarianship, on one side, and Arab studies and research on the other side. The "Libraries of the Middle East" project aims to enhance the visibility and dissemination of those holdings of the BnF relating to the Middle East together with those of partner heritage and research libraries and archives, most of them located in the Middle East. As such, the project could play the role of a platform to improve exchanges between libraries and researchers and mobilise the resources (financial, scientific, and technological) necessary to ensure a better digital visibility and accessibility for collections of Arabic literature. This presentation would like to kick start a conversation on what type of mechanism could be put in place to improve communication and pool resources between two complementary professional domains in view of meeting more adequately the expectations of scholars as regards digital access of Arabic manuscripts while taking into account the operational constraints the BnF and its partners work with.

Bushra Farah

Sayf bin Dī Yazan en ligne : Réflexion sur les méthodes de collecte, de classement et d'analyse d'un grand corpus de manuscrits numérisés.

Cette communication interroge un travail en cours portant sur le roman populaire *Sayf bin Dī Yazan*. Elle se propose comme une réflexion méthodologique sur une opération de collecte de documents qui a duré près d'un an ainsi que sur les problèmes et obstacles rencontrés. En effet, cette opération m'a permis de rassembler une cinquantaine de volumes de *Sīrat Sayf* appartenant à une vingtaine de manuscrits, complets ou fragmentaires. Il est à ce stade difficile de me prononcer sur la pertinence du résultat de cette collecte pour atteindre l'objectif premier de mon étude sur cette *sīra*. Cet objectif est celui d'étudier la naissance et l'évolution de la version damascène à partir de la version égyptienne. Or, comme il a été déjà observé pour d'autres *siyar* comparables, établir un stemma dans le sens classique du terme est une mission qui paraît à l'heure

actuelle longue et peut-être pas suffisamment rentable, avec un échantillon comme le nôtre, dont nous vous exposerons les caractéristiques. Et ceci n'est pas étonnant : un roman qui est né probablement dans le courant du XIV^e siècle et qui a été transmis en partie oralement et à travers de très nombreuses copies au moins jusqu'au milieu du XIX^e siècle, simultanément dans différentes régions, doit avoir des centaines de versions. Notre échantillon n'est qu'un petit extrait d'une « galaxie » (pour reprendre le terme de Herzog) de versions, qui peut difficilement nous fournir des réponses sur la genèse du récit. Toutefois, l'état hétérogène de notre échantillon met en évidence une palette de procédés de production des manuscrits qu'il serait intéressant d'examiner, et qui pourrait éventuellement constituer une piste à suivre dans l'étude de l'évolution du récit. Autrement dit, si les données codicologiques et linguistiques se joignent aux données textuelles dans l'étude d'un petit échantillon de manuscrits, jusqu'où pourrait-on aller dans la restitution de la genèse du récit lui-même ? Il s'agit, dans le cadre de cet atelier, de présenter une ébauche de réflexion sur une méthode d'approche adaptée à ce corpus : une méthode que nous pourrions développer au cours de notre travail de thèse, suivant les premiers résultats obtenus. Autrement dit, nous expérimentons la pertinence d'une manière d'aborder un corpus assemblé, comme le sont souvent les corpus de cette littérature arabe dite « populaire », au hasard des circonstances souvent peu commodes.

Sayf bin Dhī Yazan online: Reflection on the methods of gathering, classifying and analysing a large corpus of digitized manuscripts

The communication is about a work in progress on the popular epopee of *Sīrat Sayf bin Dhī Yazan*. It is a reflection on a document collect operation that lasted almost a year: the encountered difficulties and unexpected results. This operation allowed me to collect about fifty volumes of *Sīrat Sayf* belonging to about twenty manuscripts, complete or fragmentary.

It is still difficult to appreciate the relevance of the resulting collection in achieving the main objective of my study on this *sīra*, which is to trace the evolution of the Damascene version starting from the Egyptian one. However, as has already been observed for other comparable *siyar*, establishing a stemma in the classic sense of the term is a mission that currently seems long and not sufficiently profitable, with a sample like ours, whose characteristics will be exposed. Is this surprising?! An epic which was probably born in the 14th century, and which was transmitted partly orally and through many copies at least until the middle of the 19th century, simultaneously in different regions, must have hundreds of versions. Our sample is only a small extract from a "galaxy" (to use Herzog's term) of versions, which can hardly provide us with answers about the genesis of the story. However, the heterogeneous state of our sample highlights a range of manuscript production processes that would be interesting to examine, and which could possibly constitute a path to follow in the study of the evolution of the tale. In other words, if codicological and linguistic data join textual data in the study of a small sample of manuscripts, how far could we go in restoring the genesis of the story itself? Within the framework of this workshop, we aim to present a first simple reflection on an approach that might seem to be adapted to this corpus, a method that we could develop during our thesis work. In other words, we are experimenting the relevance of a way of studying a corpus gathered according to the circumstances, often complicated, as it goes for this kind of corpuses of the so-called "popular" Arabic literature.

Zuzana Gasakova

Traits caractéristiques des manuscrits de Sīrat Sayf ibn Dhī Yazan

Les manuscrits de *siyar sha'biyya* (sg. *sīra*) représentent les sources les plus authentiques d'épopées populaires arabes qui nous soient accessibles de nos jours, car le nombre d'enregistrements est faible, et témoignent malheureusement du statut amoindri de la tradition narrative orale au XX^e siècle. La majorité de ces textes n'a pas fait l'objet d'un examen approfondi

pour diverses raisons, mais ils méritent néanmoins une attention particulière de la part des chercheurs.

Dans la contribution suivante, je voudrais me concentrer sur ces documents volumineux en m'appuyant sur l'important corpus de *Sīrat Sayf ibn Dī Yazan* (par exemple Le Caire : Dār al-Kutub, MS n° 4592 ; MS n° 13524 ; MS n° 91 ; Leipzig : Bibliothèque universitaire MS K. Vollers 630 ; Leyde : Bibliothèque universitaire de Leyde MS 1680 a, b ; Berlin, Bibliothèque d'État MS We 643 ; Londres, British Library, MS Or 4643 ; Milan, Bibliothèque Ambrosienne, MS n° 148 ; Paris , Bibliothèque Nationale de France MS n° 3812 et quelques autres). Les manuscrits d'autres *siyar* (*Sīrat Dhāt al-Himma*, *Sīrat Badrnār ibn Nahrawān*, etc.) seront également consultés pour démontrer certaines particularités.

Mon attention se concentrera sur des caractéristiques visuelles spécifiques de manuscrits individuels comme l'aspect général des « livrets » (pourquoi en avons-nous tant dans certains cas ?), une page de titre, l'écriture (plusieurs mains, type d'écriture), l'ouverture de la *sīra*, la vocalisation, la subdivision du texte, les colophons, la mise en page, etc. Ce faisant, je commenterai également les compétences linguistiques du scribe par rapport au genre, les techniques de composition et les divergences dans la trame de la *sīra*, qui confirment le caractère oral du texte. En raison de l'oralité, les conteurs ou les scribes pouvaient aussi facilement agrandir ou raccourcir tout épisode de la vie du héros. Un tel épisode pouvait ainsi être élaboré de manière détaillée en prose ou parfois en vers ou, au contraire, considérablement raccourci, laissé sans développement, mentionné au hasard ou complètement omis. Il faut également mentionner la négligence occasionnelle du scribe, par exemple lorsqu'une partie considérable du texte a disparu au cours du processus de copie. Il est également possible de faire la distinction entre les manuscrits qui ont été utilisés, très probablement, comme support d'exécution (ou comme un certain aide-mémoire) et ceux qui ont été compilés pour un usage privé.

Bien qu'il n'y ait pas de cas de formulation identique de grandes parties des manuscrits de *Sīrat Sayf*, même si la formulation est souvent très similaire, il est possible de reconnaître certaines relations et parallèles entre ces textes qui peuvent jeter quelques lumières sur la façon dont la narration s'est progressivement développée. À cet égard, il est possible de distinguer au moins deux « stemmas » particuliers de la narration qui, avec la typologie des changements, seront également discutés dans ma contribution.

Il va sans dire que la présentation sera accompagnée des images des folios des manuscrits en question.

Characteristic Features of the Manuscripts of Sīrat Sayf ibn Dhī Yazan

Manuscripts of *siyar sha'bīya* (sg. *sīra*) represent the most authentic sources of Arabic popular epics accessible to us nowadays as the number of recordings is low and unfortunately bear witness to the diminished status of the oral narrative tradition in the 20th century. The majority of these texts has not been thoroughly examined for various reasons, however, they are well worth scholarly attention.

In the following contribution, I would like to focus on these voluminous materials by relying on the large corpus of *Sīrat Sayf ibn Dhī Yazan* (e.g. Cairo: Dār al-Kutub: MS N° 4592; MS N° 13524; MS N° 91; Leipzig: University Library MS K. Vollers 630; Leiden: Leiden University Library MS 1680 a, b; Berlin: State Library MS We 643; London: the British Library, MS Or 4643; Milan: Ambrosiana Library, MS N° 148; Paris: Bibliothèque Nationale de France MS. N° 3812 and some others). The manuscripts of other *siyar* (*Sīrat Dhāt al-Himma*, *Sīrat Badrnār ibn Nahrawān*, etc.) will also be consulted to demonstrate some particular features.

My attention will be concentrated on specific visual features of individual manuscripts as the general appearance of "booklets" (why in some cases do we have so many of them?), a title page, script (several hands, type of script), the opening of the *sīra*, vocalization, subdivision of the text, colophons, layout, etc.). In doing so I will also comment upon the scribe's language competences

in relation to the genre, compositional techniques and discrepancies in the *sīra*'s plot, which confirm the oral nature of the text. Due to the orality, storytellers or scribes could also easily enlarge or shorten any episode from the hero's life. Such an episode could be thus, on one hand, elaborated in a detailed manner in prose or sometimes in verses or, on the other hand, considerably shortened, left undeveloped, randomly mentioned or completely omitted. The scribe's occasional negligence must also be mentioned – e.g. when a considerable part of the text somehow disappeared during the process of copying. It is also possible to distinguish between manuscripts which were used most probably as performance prompts (or as a certain *aide-mémoire*) – and those compiled for private use.

Despite the fact that there are no cases of identical wording of extensive parts of manuscripts of *Sīrat Sayf*, even though the wording is often very similar, it is possible to recognize certain relationships and parallels between these texts which may cast some lights on how the narration gradually developed. In this respect, it is possible to distinguish at least two particular "stemmas" of the narration which alongside with typology of changes will also be discussed in my contribution.

It goes without saying that the presentation will be accompanied by the images of particular manuscript folios in question.

Iyas Hassan

Vers un modèle collectif de l'édition critique des manuscrits. Réflexion sur le projet Alexandre (2022-2025)

Développé dans le cadre du programme ANR - « Littératures Populaires du Levant » (LiPoL), le projet Alexandre (2022-2025) a pour ambition de faire évoluer le modèle d'édition critique développé auparavant dans le projet Baybars. En effet, la disponibilité des manuscrits numérisés de la tradition populaire arabe du Roman d'Alexandre (*Sīrat al-Iskandar*) et la possibilité désormais de recourir aux techniques HTR pour produire des données textuelles massives à partir des manuscrits, contrastent avec les conditions matérielles difficiles dans lesquelles a été réalisée l'ambitieuse édition de *Sīrat al-Malik al-Zāhir Baybars* (2000-2022 ; 18 vol. ; Bohas, Diab, Hassan, Zakharia). Le matériel et la technologie à disposition aujourd'hui imposent de *facto* une réflexion sur le mode opératoire et sur la nécessité de croiser les enjeux scientifiques (l'édition à proprement parler) à ceux techniques (développement d'un modèle HTR) et de formation à la recherche et par la recherche (place des étudiant·e·s dans ce processus).

Cette communication consistera en un retour d'expérience sur le Hackathon Alexandre (décembre 2022-mai 2023), première étape du projet Alexandre, et en une mise en perspective des pratiques que suggère cette nouvelle expérience dans ce nouveau contexte, en matière d'approche des manuscrits et d'édition critique.

Towards a collective model of the critical edition: Reflections on the Alexandre Project (2022-2025)

Developed as part of the "Littératures Populaires du Levant" program (LiPoL), the Alexandre Project (2022-2025) aims to develop the critical edition model previously initiated in the Baybars Project. The availability of digitized manuscripts from the popular Arabic tradition of the Alexander Bio-epic (*Sīrat al-Iskandar*) and the use of Handwritten Text Recognition (HTR) techniques to produce massive textual data from the manuscripts, is in contrast with the difficult material conditions in which the ambitious edition of *Sīrat al-Malik al-Zāhir Baybars* (2000-2022; 18 vol.; Bohas, Diab, Hassan, Zakharia) was produced. The new context in which this work is being done imposes a number of considerations we must reflect on, especially the intersection of the technical, the scientific and the pedagogical aspect. This talk will consist of feedback on the Alexandre Hackathon (December 2022–May 2023), the first stage of the Alexandre Project, and of putting into perspective the practices suggested by this new experience in this new context, in terms of approaching manuscripts and critical editing.