Faculté des langues
Université de Strasbourg

Groupe d'études orientales, slaves et néo-helléniques | UR 1340

Université de Strasbourg
Initiative d'excellence

(Département d'**études arabes**)))

Le projet HUNAI (Humanités Numériques pour l'histoire de l'Arabie Islamique)

Eric Vallet et Clarck Junior Membourou Moimecheme

L'HTR des langues peu dotées dans les programmes de recherche et dans les établissements de conservation français – 14 février 2024

UAR2999 distam

BnF

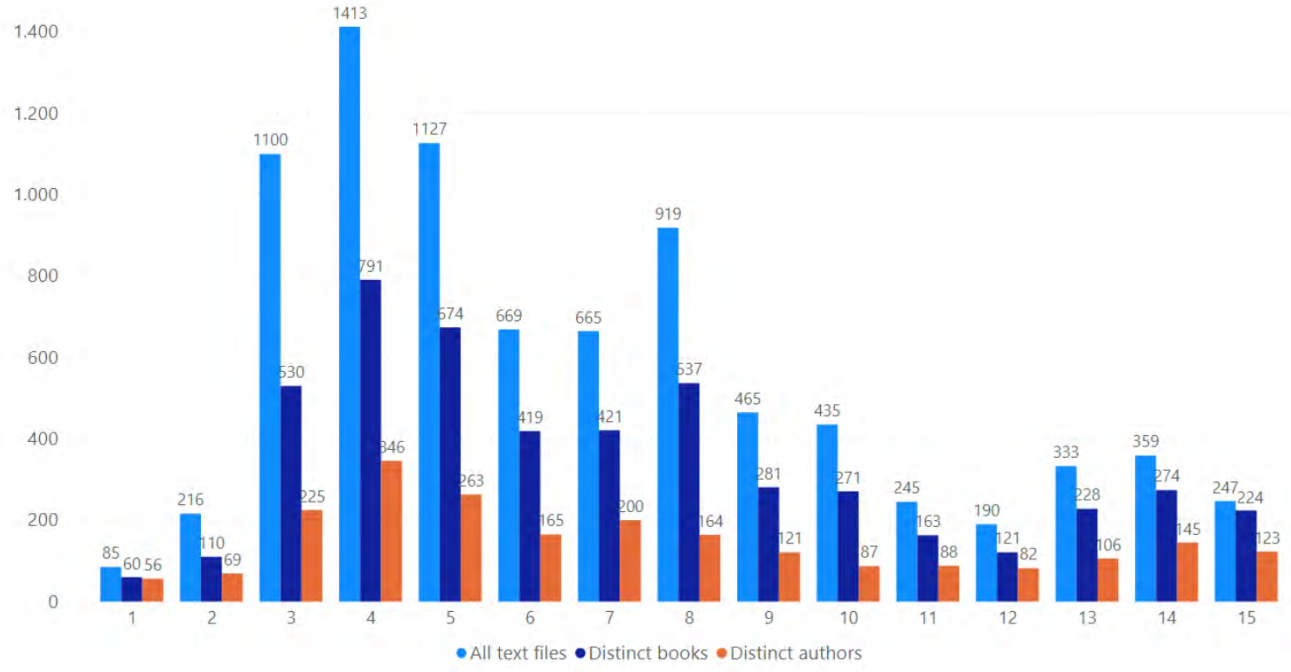
arabes

- **6 millions de manuscrits islamiques dans le monde ?**

7280 manuscrits arabes à la BnF

- **La “Constellation” OpenITI**

KITAB Project > Open Islamicate Text Initiative OpenITI (Maxim Romanov, Sarah Savant and Matthew Miller)
10 202 fichiers textes ; 6 236 “oeuvres” de 2 582 auteurs (dernière livraison sur Zenodo).



Text files in OpenITI, per hijrī century (light blue: number of text files; dark blue: number of distinct books; red: number of distinct authors). NB: pre-Islamic texts are in the first-century column, texts by authors still alive in the 15th-C column. (data: Dec. 2020)

Source :
https://openiti.org/documentation/#faq

L'HTR et manuscrits en écritures arabes

Benjamin Kiessling, *Avancées en Reconnaissance Optique des Caractères pour les Documents Arabes Historiques,* thèse EPHE 2021.

***Open Islamicate Texts Initiative Arabic-script OCR Catalyst Project (OpenITI AOCP)* 1 et 2 (2019-2021)**
escriptorium

***Automatic Collation for Diversifying Corpora* (*ACDC*) (Matthew Miller) 2024-** – Financement National Endowment for the Humanities 2024 = Automatisation de l'entraînement de modèles HTR sur manuscrits arabes et persans

- Combinaison de Kraken, de l'outil Passim (text reuse) et des textes numérisés dans Open ITI (alignement automatique des textes issus de l'HTR à partir des textes déjà numérisés dans Open ITI)

https://github.com/OpenITI/ACDC
https://www.youtube.com/watch?v=kNx4GyH5HSo

Littératures populaires arabes

Hackaton Alexandre (2023)

https://gitlab.huma-num.fr/lipa/iskandar

2. HUNAI : genèse et objectifs

➢ Genèse

- Une vision étriquée de l'Arabie, terre sans histoire
- Un patrimoine manuscrit partiellement sauvegardé et valorisé par les institutions locales ou internationales
- Une production textuelle peu considérée, sous-exploitée et sous représentée

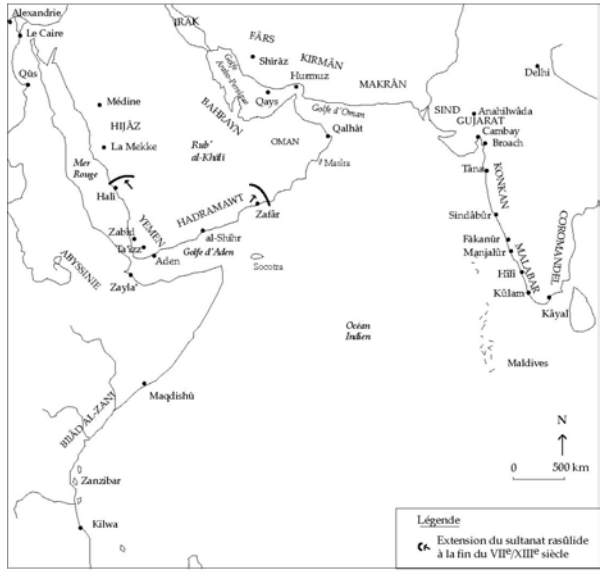
➢ Objectifs

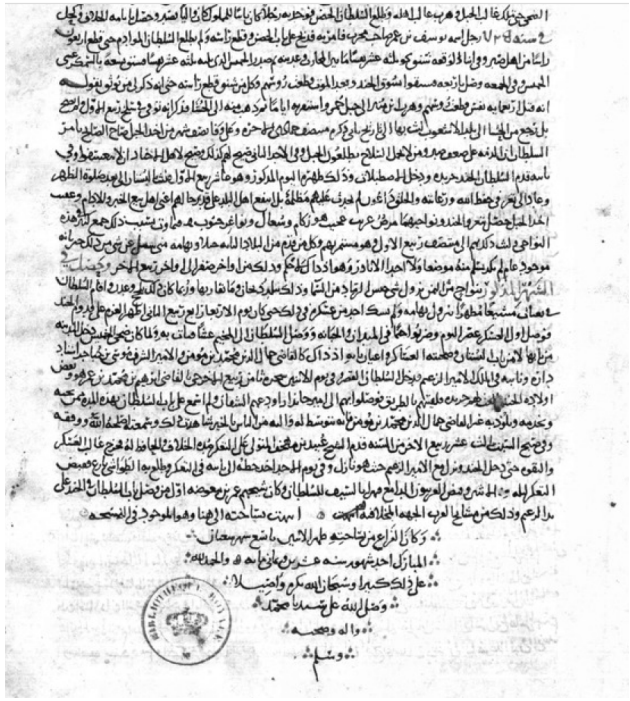
- Réintégrer un corpus cohérent d'ouvrages majeurs **dans l'ensemble des ressources numériques arabes disponibles en *open access*** et annotées selon les standards internationaux
- Développer des **études et des analyses historiques et linguistiques reposant sur la fouille de données textuelles** appliquée à des corpus cohérents issus de cette production.

Corpus

- Composé de données textuelles correspondant aux ouvrages historiographiques sunnites composés au Yémen à l'époque rasūlide (626-858/1229-1454)

- 6 manuscrits numérisés de la Bnf, publiés sur Gallica
  - **al-Janadī, *Sulūk fī ṭabaqāt al-'ulamā' wa-l-mulūk*, BnF Arabe 2127 (copié en 820/1417)**
  - **al-Ashraf Ismā'īl/al-Khazrağī, *Al-Kifāya wa al-i'lām fī man waliya al-Yaman wa sakanahā fī al-islām*, Paris BnF 5831**
- 9 mss (Yale ; Institut des manuscrits arabes du Caire)





BnF 2127 – colophon du *Kitâb al-sulûk* d'al-Janadî

3. L’annotation : genèse, méthodologie et adaptations

3.1. Genèse de l’annotation

1 modèle arabe générique plus récent

1 لسعلمفم الله الرحمن الرحيم ف و الحمد لله الملك العطام الاول الاحر العدلم مما عب نسه
2 محميدها عرلا الحليل والتحريم وداعنا الى الصراطا المسبتعيم قبلع صلى الله عليه وسلم الرساله على النعهم وهم لمن تعه
3 الدين القوهم مو عن على الكافه بصديقه و العام يواحب الصلوة لنه والتسليم فله الحمد حمد الحرج عن درك الاخصاه
4 ويدغم به طايف من حمد وعسره وصلى الله على سة المنعوث صلوه استاسرها الحنن ويحشه القتور واتميربها في الدارس

2 modèle arabe générique ancien

1 سمليقم اننه الحمن الركم و الحمديته المك العطم الاول الاحر العدكم ما عب نسه
2 محمد مهمامه با الحليل والنحريم وداعنا الى الصاط المستقيم فبلع صلى الله عليه وسلم الرسالة على النعهم ونهم ان تعه
3 الدين القويم فوع على الكافه يصديقه والقام يواجب الصلوة طيه والتسليم ه فله الحمد حمد الخرج عن دركالاحصا
4 ويرغم به طرايف من حمد وعسي وصلى الله على مسة المبعون صلو استاس هالحين او يحثه القبور كواتمرتها في الدارس

3 modèle Hamawi Strasbourg

1 لسمايفه االخمر الرجي والهية الك الحطم الولا الجر اليهع ماعب سه
2 محبرها لمرلا الحليل والنحرم ودا عبا الي الطرا المستعيم معبلع ملى ابه عليه وسلم الرباله علي النغميم موتهم ل تعه
3 الدسن العوه حمو ملن الكافه يضويقه والعيام يواج الطلواطبه والسشلم فله اهحمد ا يحرج عن در اا حصا
4 ويرغميه انف من حد وعسي وطى اه لي سة المبعوث صلوه استاس ماحتن ويحشه التور توا تميربا في الرارر

Annotations du f° 62 r. du BnF arabe 5832 dans un fichier .txt

فزحفت العساكر المنصورة على الحصن ثلاثة أيام متوالية وكتب الأمير علي ابن عبد الله إلى كافة الأشراف كتبا متتابعة يطلب منهم النصرة وهم يغالطونه ويعتذرون ثم حصل خطاب ومراسلات علي يعنى الصلح
واستقر الحال على أن الأمير علي بن عبد الله يواجه الصاحب موفق الدين فوصل إليه. واتفق حضور الملك المظفر فاجتمعوا جميعا وساروا بأجمعهم إلى المقام الشريف السلطاني فلما علم السلطان بوصولهم ركب من مخيمه وقد صاروا بالقرب منه وأكرمه وانصفه وانعقد الصلح بينهم وأخذ للأشراف ذمة سبعة اشهر
وسلم لأجلها حصن ذيفان لان السلطان امتنع من الذمة عليهم فلما استقر في المحطة طلب من السلطان دخول الإعلام الشريفة اظهرا اللطاعة والتسليم فنصبت في أعلا الحصن وكذلك العظيمة فخفقت ذوائبها في أعالي الحصنين ولقد أحسن الحسن بن هاني يقول من كان بالسمر العوالي خاطبا ... جليت له بيض الحصون عرائسا
و قال العفيف عبد الله ابن جعفر يمدح السلطان الملك المؤيد ويذكر أخذه العظيمة واللميفاع

إرث الخلافة في يديك مشاع ... وعرار سيفك شاهد قطاع
منع النصيب من العد انصب القان ... والجم القراع من اليوف قراع
شمس رأت غلب الملوك شعاعها ... فقلوبها منها تطير شعاع
تبع التتابع في عناصر حمير ... والي المناقب هم له أتباع
عمرو وعمرو ذو الجناح ومنذر ... والأيمان وفايش وكلاع
ماء السماء سقى منابت اصله ... ربا فأورق عرقه النزاع
فلقد أعاض بيوسف يقطان فلا ... نكل ولا كل ولا مجزاع
أشرا إلى الشرق القصي بشرب ... خطواتها نحو المراع سراع
والشمس من لمع الحديد كليلة ... والجو من سمر اليراع يراع
وفيالق سالت هوادئ خيلها ... سيل الأتي تداولته تلاع
تسري فمن زرق الأسنة فوقها ... نار ومن أسل الوشيج شعاع
غسلت مياه سيوفها ماء الدجي ... فتشابه الإصباح والأهزاع
ينحو بها مبدأ النجوم طوالعا ... ملك مطيع للإله مطاع
ليس العظيمة بالعظيمة عند من ... بسيوفه ميقاعها ميقاع
لم يشق وافدهم إليه اتي وهل ... يشقى أمره وجليسه القعقاع
فغنمت أدعية بأفواه لهم ... فيهن من ثدي البتول رضاع
وحفظت حقا للنبي محمد ... فيهم ولست بما حفظت تضاع
أيام الإسلام سيف وضع ... الوسر مجك من الشاسطاع
أمؤيد الإسلام داود الذي ... للعالمين بفضله إجماع
ما يلتقي شرق البلاد وغربها ... إلا إذا ما امتد منك الباع
أهويت بالسيف العداة كما هوى ... ود بسيف محمد وسواع
الله أعطاك السعادة كلها ... ماذا يضر وربك النفاع

وهي أطول مما ذكرت وهذه عيونها و اقبل السلطان على الأمير جمال الدين علي ابن عبد الله المحبة وأرال ما في خاطره وجدد له رفع الطبلخانة وحمل معها من الكساوي والاموال شيئا كثيرا ولما كان أول يوم من شهر ربيع الأول ارتفع السلطان من محطة قاضى صنعاء
أمام الكتيبة تزهى به ... مكان السنان من العامل
قال الشريف إدريس وسرت في خدمته مع والدي الي الوعدت من هناك وقد كنت خرجت

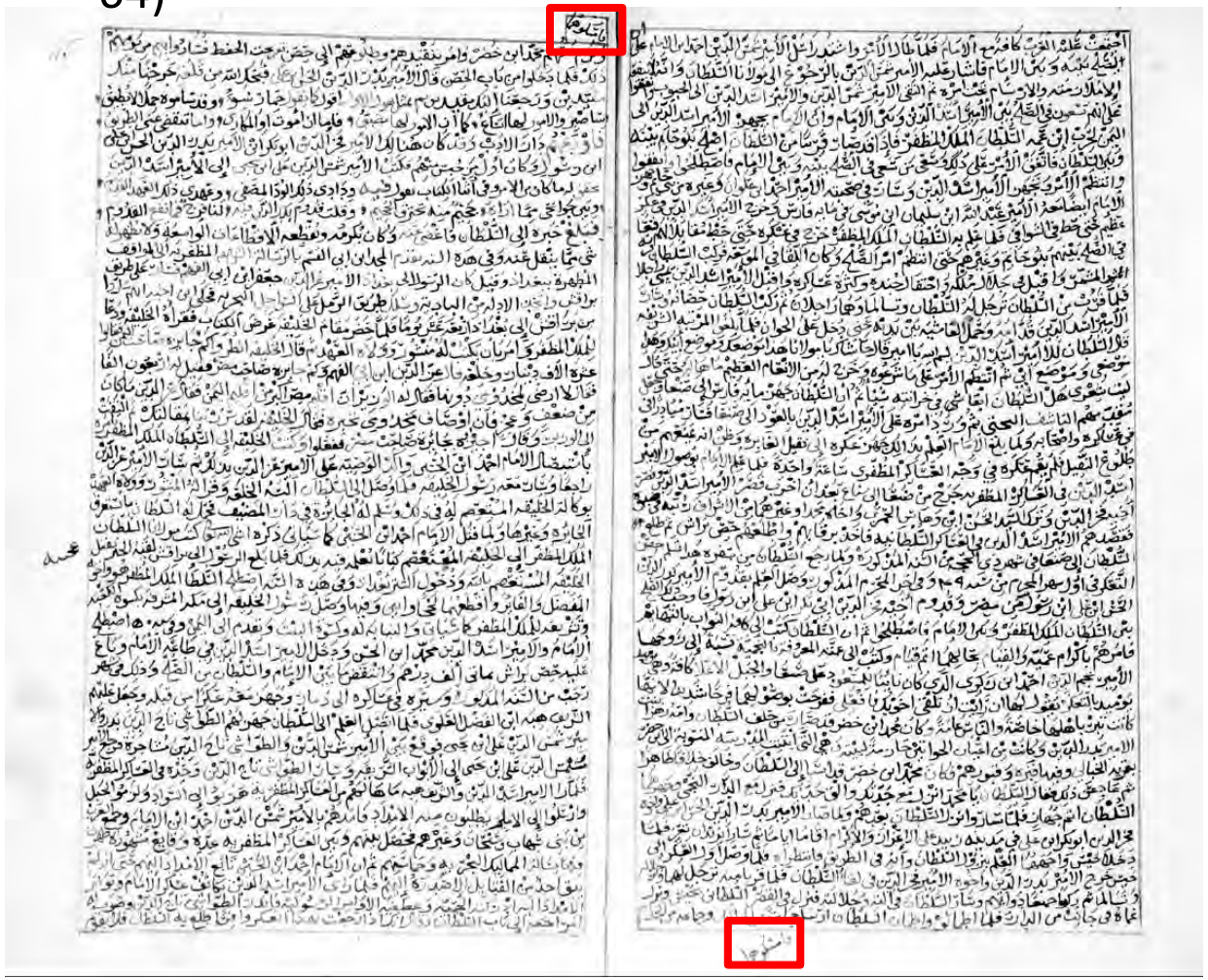
Annotations du f° 62 du BnF arabe 5832 importées dans le logiciel Calfa

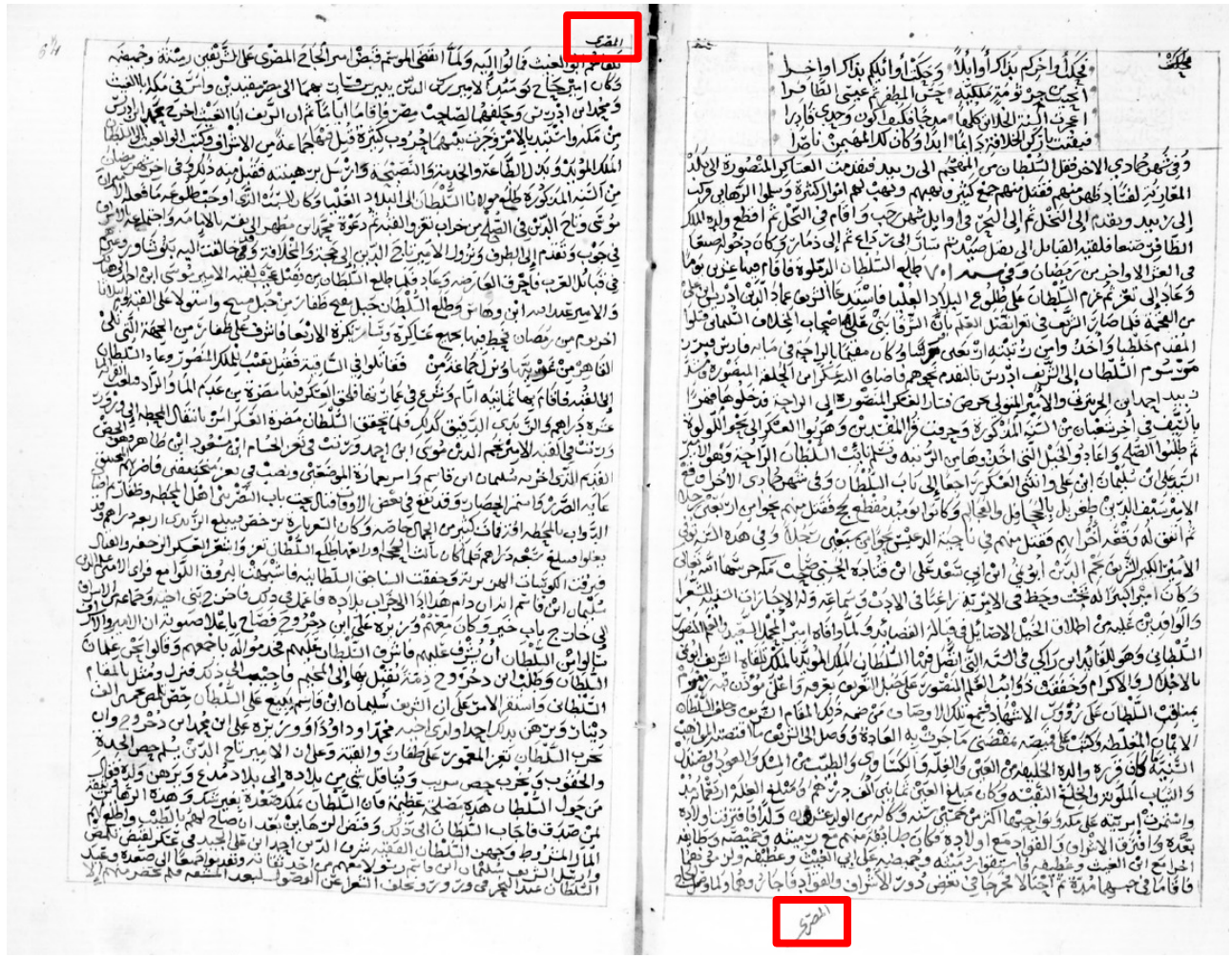
3.3. Adaptations

Réclames, notes marginales et entre les lignes

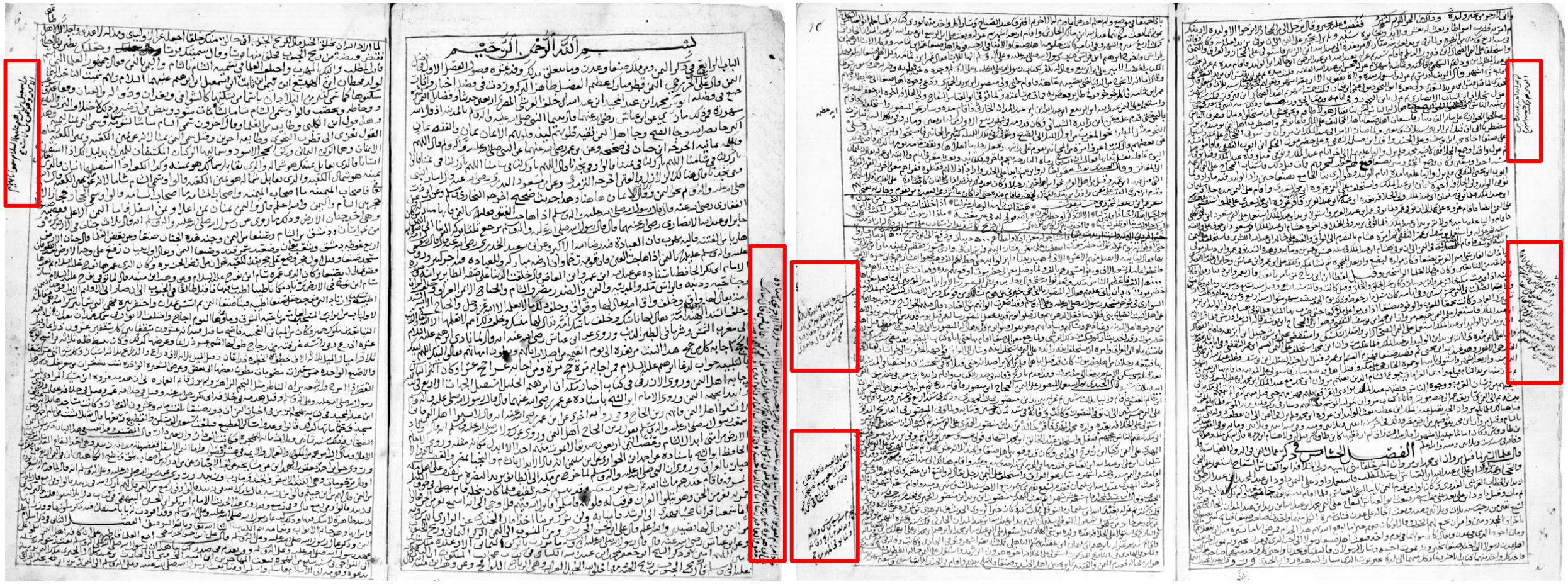
La réclame est le terme placé en bas d'une page manuscrite et qui correspond au premier terme de la page suivante.

Calfa considère le premier mot de la page suivante comme une note marginale. (Cf. BnF arabe 5832 f° 45 et 64)

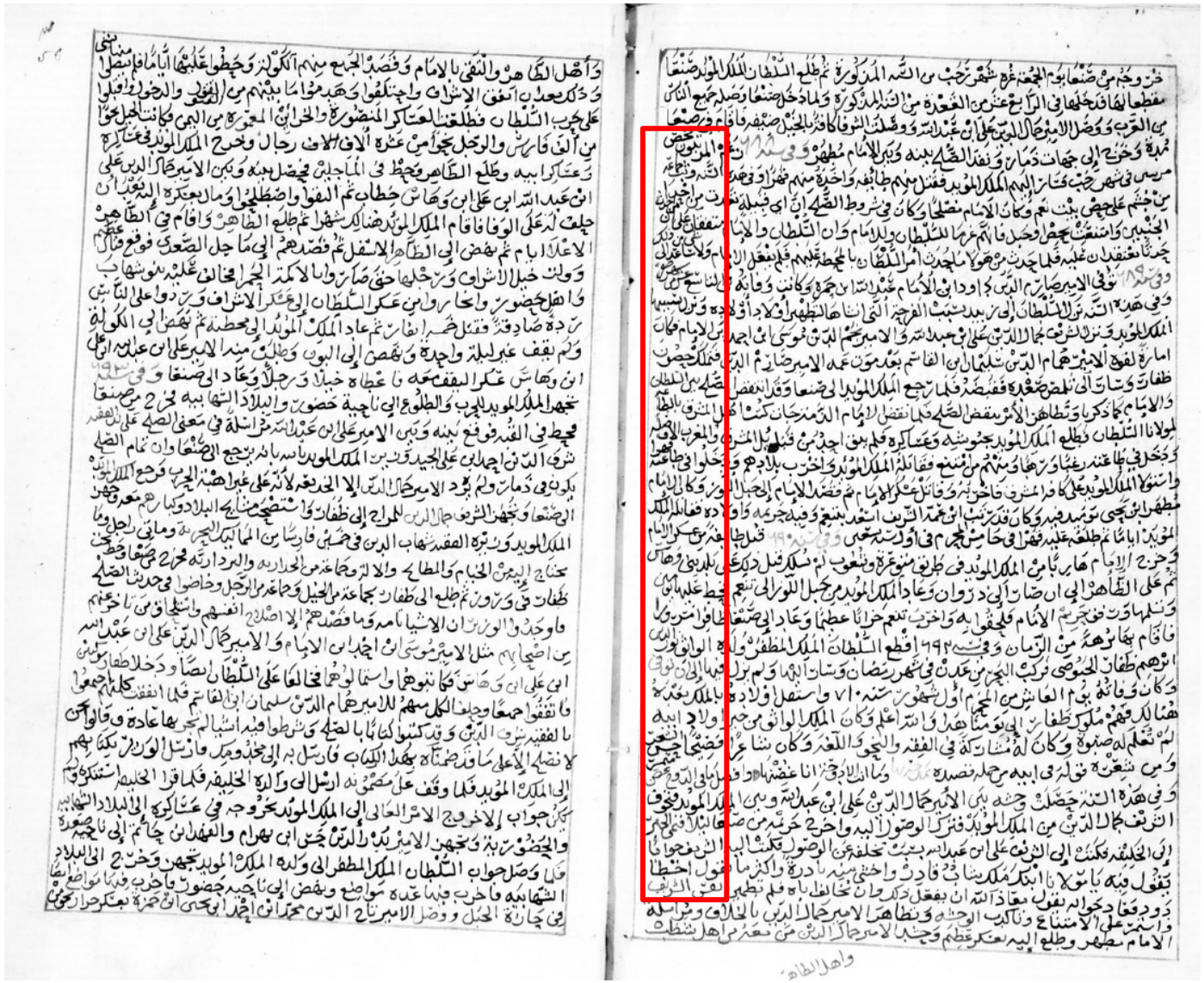




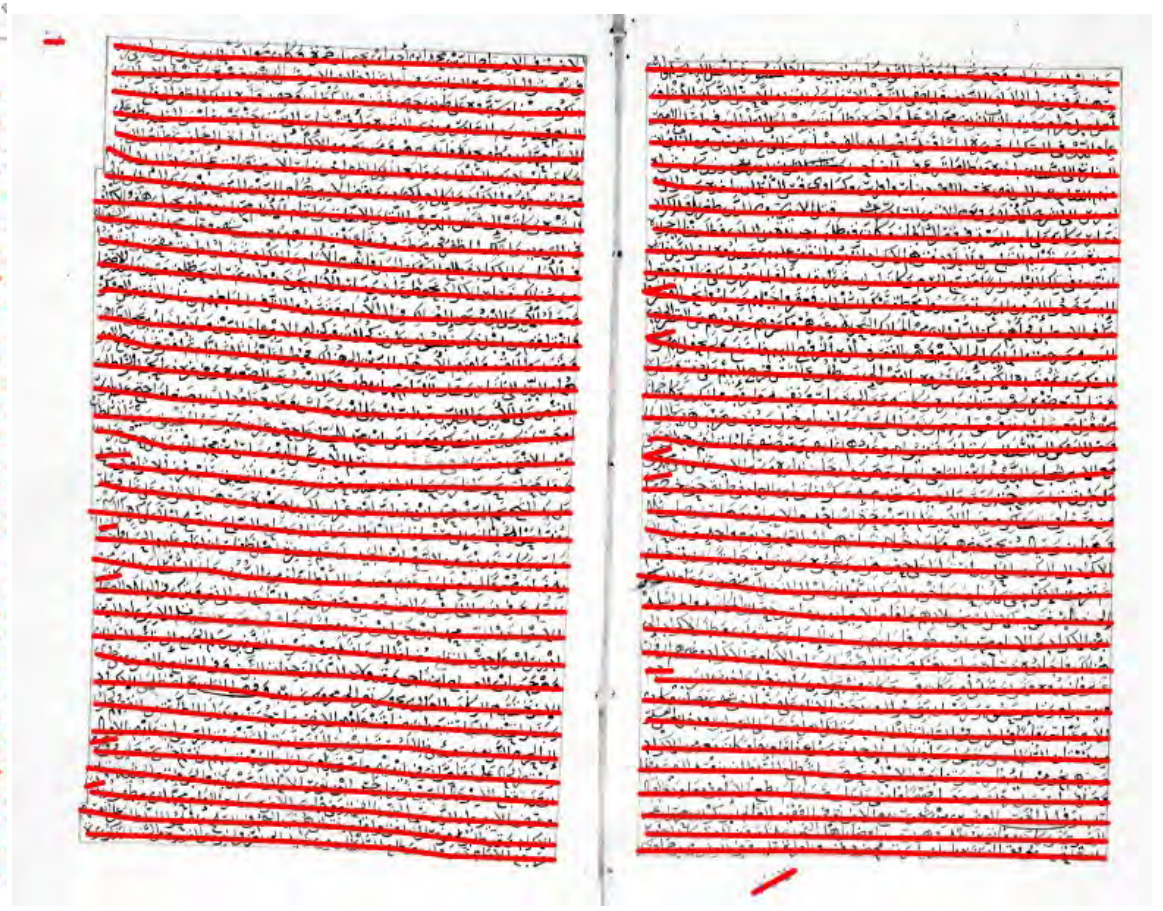
Présence importante de notes marginales dans les premières pages du mss BnF arabe 5832 (ex : f°6 et 10)



L’ajout régulier des termes entre les lignes rend la lecture et l’annotation difficiles



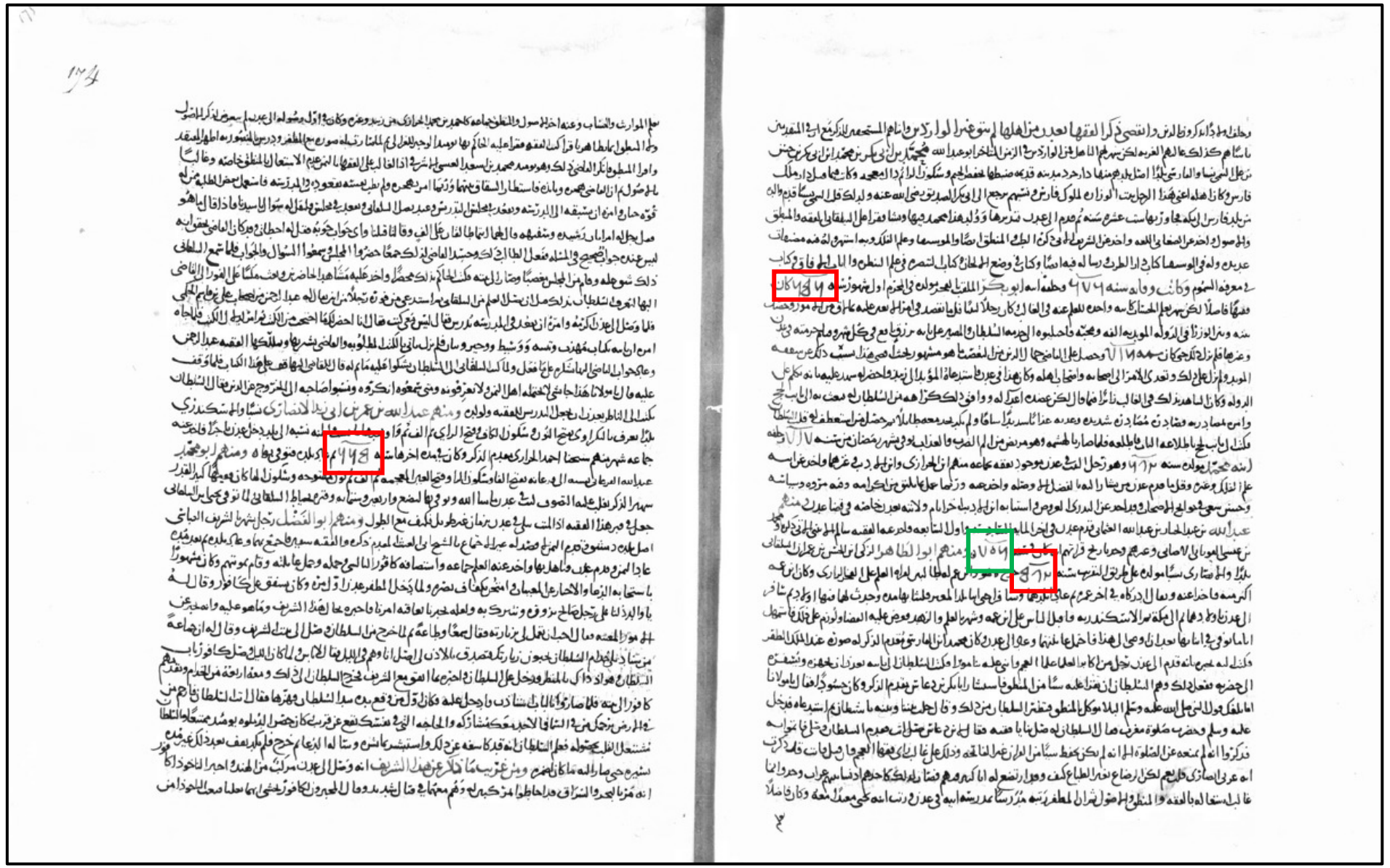
Les lignes de base appliquées sur les mss BnF arabe 2127 et 5832



Les chiffres et les lettres

Les dates apparaissent en lettres ou en chiffres dans une même page

Le chiffre 5 est écrit de deux manières en arabe dans le BnF arabe 2127 f° 174



Utilisation irrégulière des points souscrits et suscrits obligatoires, ce qui génère des confusions entre les consonnes ب ت ث ج ح خ ص ض ي

Utilisation de ا à la place ى

Absence de la hamza ء en fin de mot

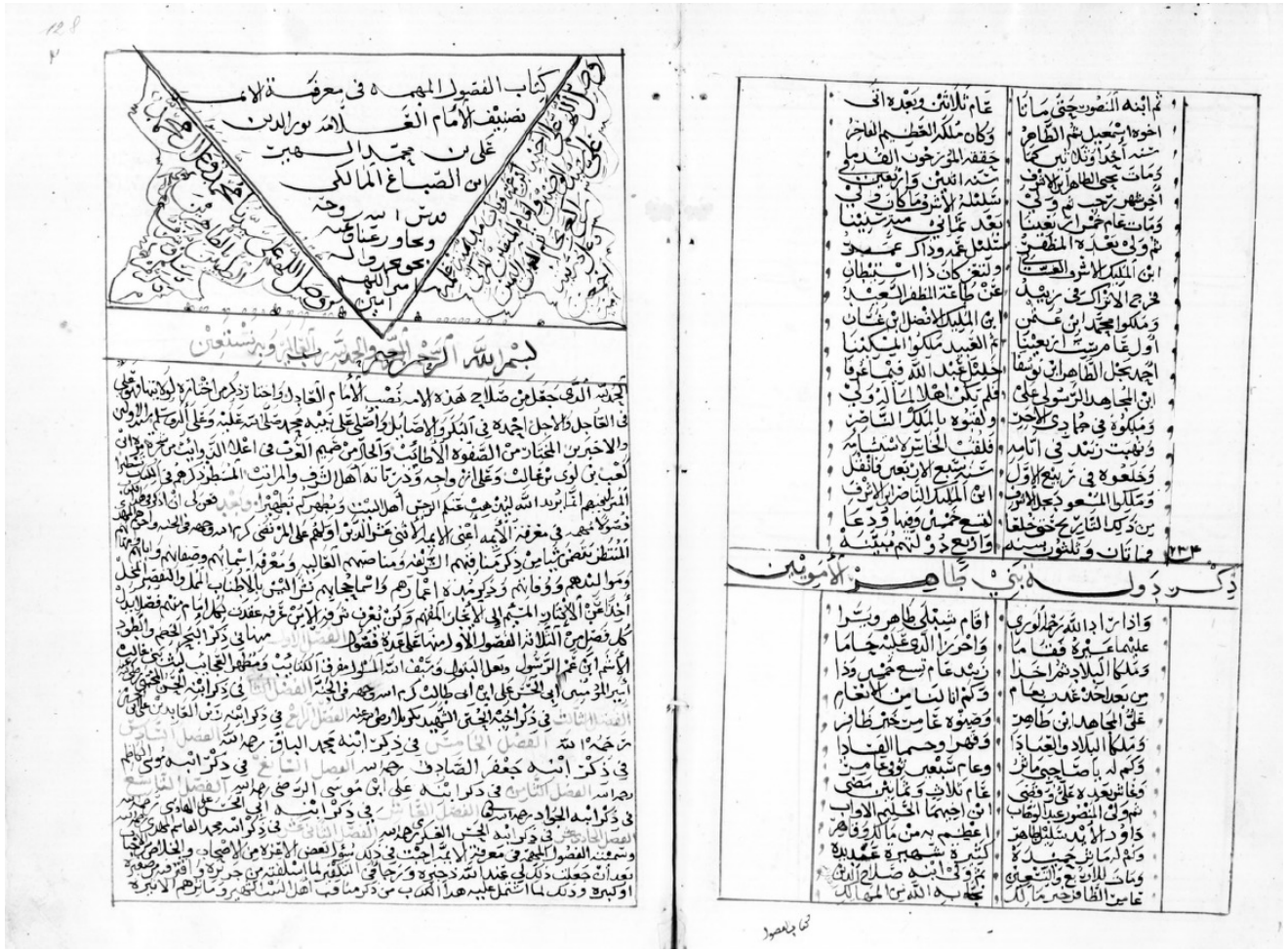
Les mots tels qu'ils sont écrits aujourd'hui	Les mots dans les manuscrits d'Hunai
علماء	علما
صنعاء	صنعا
شيء	شي

4. Un modèle adapté aux manuscrits du Yémen ?

Les annotations permettront à l'équipe de Calfa d'entraîner un modèle qui sera testé sur l'ensemble des manuscrits du projet HUNAI

Vérifications seront faites afin de déterminer le taux d'erreur, dans un premier temps sur les mss BnF arabe 2127, 5832 et peut-être 4609.

5. Des mises en page complexes à examiner avec Calfa

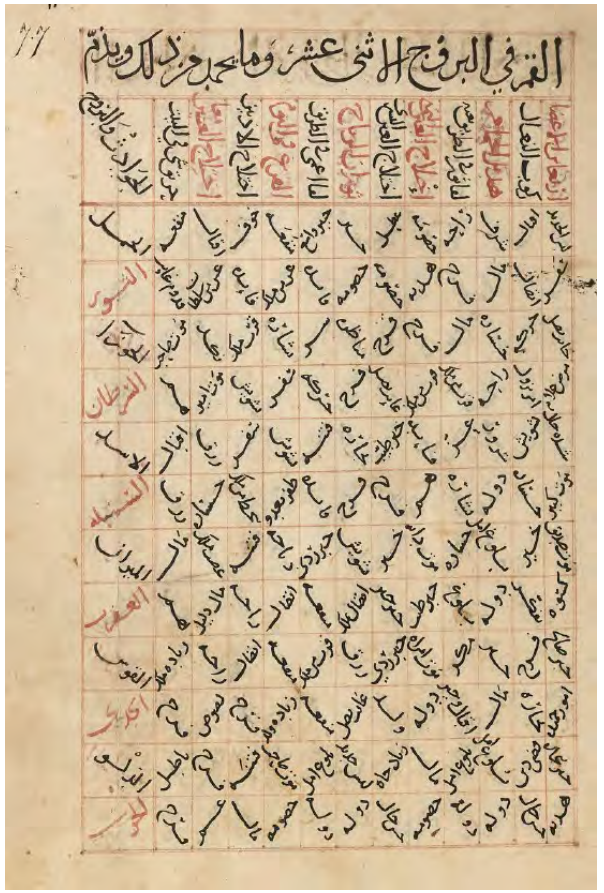


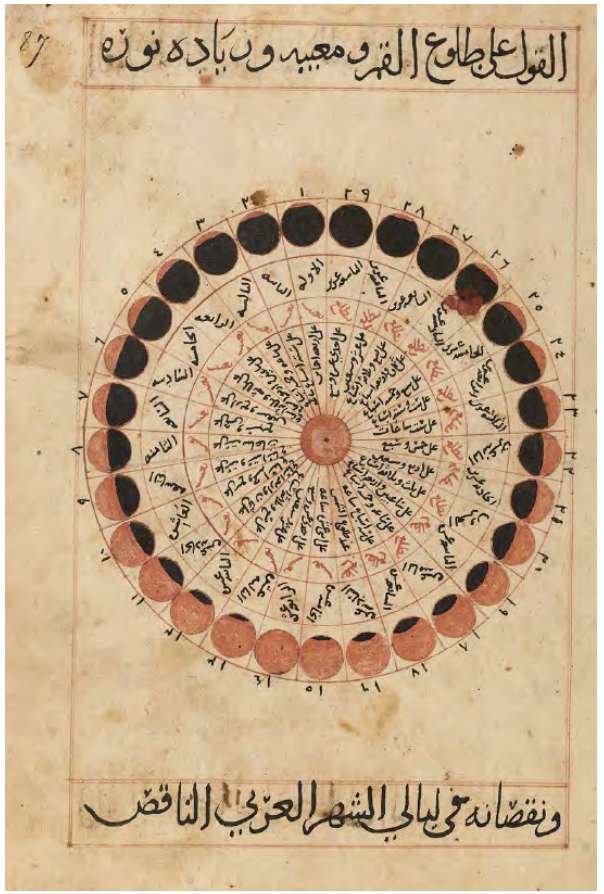
Mss BnF arabe f° 128

Comment organiser la sortie de texte ?

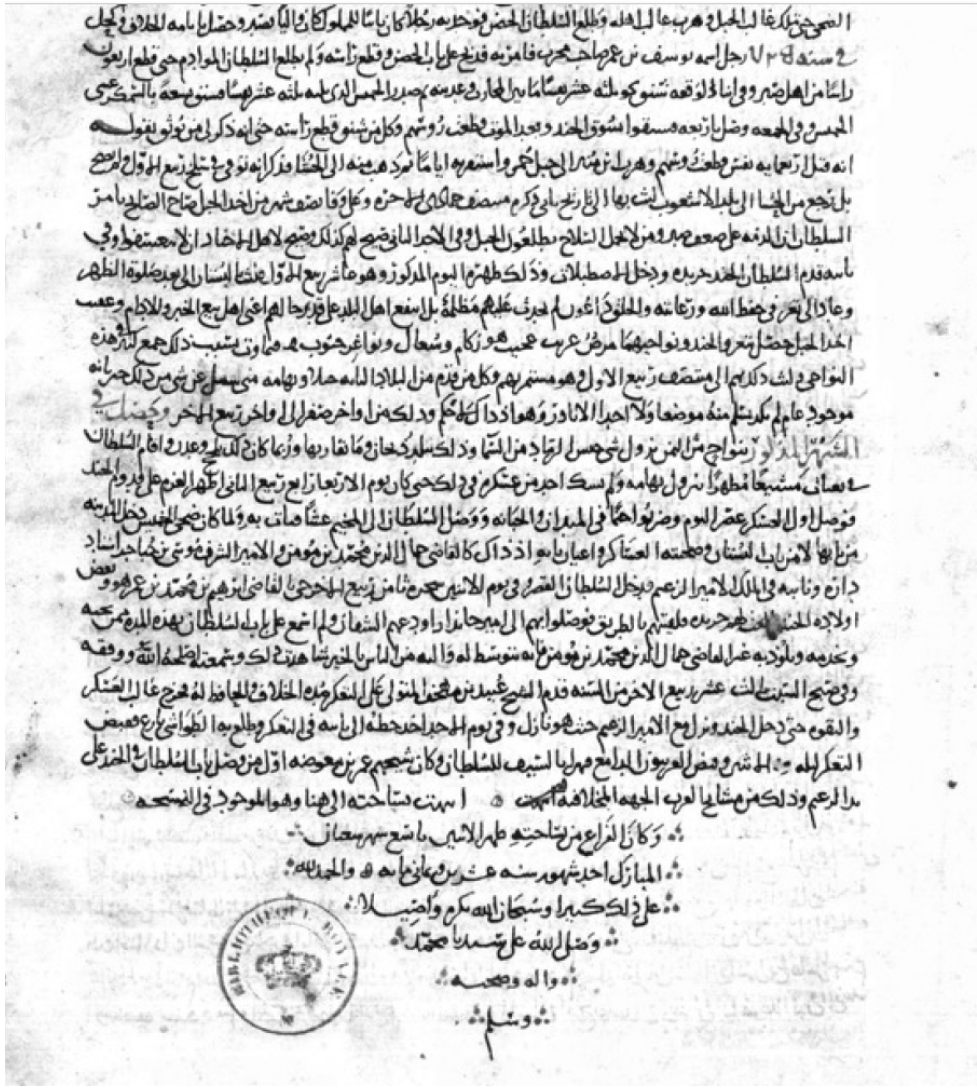
- Titres, sous-titres …
- La poésie …
- les schémas …

Mss BnF arabe 4609, f° 5, 17, 77 et 87 v.





Merci de votre attention



BnF arabe 2127 – colophon du *Kitâb al-sulûk* d'al-Janadî