



**HAL**  
open science

## Savoirs et poésie dans la Chine médiévale (CHI-KNOW-PO)

Marie Bizais-Lillig

► **To cite this version:**

Marie Bizais-Lillig. Savoirs et poésie dans la Chine médiévale (CHI-KNOW-PO). Humanités numériques en pédagogie et en recherche à la faculté des langues, Sep 2023, Strasbourg, France. Faculté des langues, Université de Strasbourg, 2024, 10.34847/nkl.f4fenvr9 . hal-04796990

**HAL Id: hal-04796990**

**<https://hal.science/hal-04796990v1>**

Submitted on 21 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SAVOIRS ET POÉSIE DANS LA CHINE MÉDIÉVALE (CHI-KNOW-PO)

Marie Bizais-Lillig et l'équipe CHI-KNOW-PO, USIAS, UR1340-GÉO, COLLEX

## Reconstruire la circulation des savoirs

- Un corpus très étendu:
  - genres élégants, ouvrages d'apprentissage, traités techniques
  - dans un effort d'exhaustivité
- Des filtres conçus pour réduire le corpus aux éléments pertinents:
  - cooccurrences
  - recouvrements et reprises séquentielles
- Un travail de lecture et d'analyse traditionnels

孟浩然 《夜泊廬江，聞故人在東寺，以詩寄之》

poem 5 of 81, no. 6703 (40 words)

江路經廬阜，**門**入虎溪。聞君尋嶽，清夜宿招提。  
石鏡山精舍，禪枝倚錦樓。一燈如悟道，為照客心迷。

孟浩然 《李少府與楊九再來》

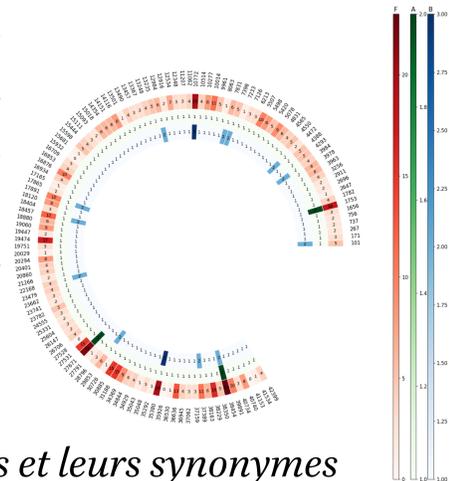
poem 6 of 81, no. 6785 (39 words)

弱歲早登龍，今來**再**逢。如何春月柳，猶憶歲寒**松**。  
**火**臨寒食，笙歌達曙鐘。喧喧門難道，行**義**朋從。

李白 《永王東巡歌十一首》

poem 7 of 81, no. 7037 (302 words)

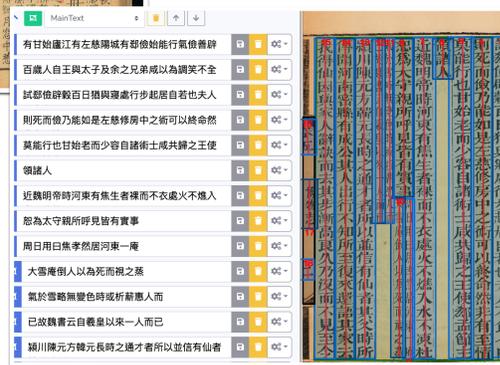
永王正月東出師，天子遙分龍虎旗。樓船一舉風波靜，江漢翻為鼉鼓池。  
三川北拂亂如麻，四海南奔似永嘉。但用東山謝安石，為君笑談胡虜沙。  
雷鼓嘈嘈喧武昌，雲旗獵獵過尋陽。秋毫不犯三吳悅，春日遙看五**色**光。  
龍蟠虎踞帝王州，帝子金陵訪古丘。春風試暖昭陽殿，明月還過鳩鵲樓。  
二帝巡**遊**俱未回，五陵**松柏**使人哀。諸侯不救河南地，更**遣**賢王遠道來。  
丹陽北固是吳關，畫出樓臺雲水間。千岩烽火連滄海，兩岸旌旗繞碧山。  
王出三山按五湖，樓船跨海次陪都。戰艦森森羅虎士，征帆一一引龍駒。  
長風掛席勢難回，海晏山傾古月摧。君看帝子浮江日，何似龍顏出岷來。



Cooccurrences de mots et leurs synonymes dans une anthologie de 50 000 poèmes



Numérisation d'un ouvrage du Collège de France et annotation sur Vision (Calfa)



## Constituer un corpus ouvert de référence

- Respecter les contraintes juridiques pour pouvoir exposer les textes étudiés
- Enrichir, structurer et rendre disponible un corpus du premier millénaire aux formats numériques (images, plein texte)
- Entraîner et partager des modèles de Handwritten Text Recognition (HTR)

### CORPUS

- Des collaborations avec les bibliothèques
- Jumeaux numériques d'ouvrages anciens
- Édition enrichie en XML-TEI

### DATABASE

- Ressource lexicale et bio-bibliographique
- MariaDB (SQL)
- Format compatible avec des bases similaires
- Liée aux corpus et scripts

### TEXT MINING SCRIPTS

- En Python (Jupyter Notebooks)
- Fouille critériée du corpus
- Sélection de textes et visualisation

## Accessibilité

Le travail est visible dans l'entrepôt GitLab de l'équipe. Chaque élément produit est déposé dans Nakala sous licence ouverte. Un site pour accéder à la bibliothèque numérique du projet est en cours de construction avec la plateforme Estrades.