

International Medieval Congress, Leeds, 2-5 July 2018

Call for Papers: *Per corpora...* Medieval Latin and Corpora (2 sessions)

Session 1: What Corpus for Medieval Studies?

Session 2: Vocabulary of Memory

Organization

Bruno Bon (*Dictionary of Medieval Latin*, Institut de Recherche et d'Histoire des Textes, CNRS)¹

Krzysztof Nowak (*Dictionary of Medieval Latin*, Institute of Polish Language, PAS)²

Motivation

Building on the experience of the session organized during the IMC 2016 we are seeking to propose two sessions for the upcoming IMC 2018 that would focus on building and using electronic corpora in Medieval Latin studies. Our idea is to stimulate discussion and integrate the community of both seasoned users and creators of digital tools, as well as those who are taking their first steps in the corpus-based research. We believe that the IMC offers a great opportunity for such discussion since the event gathers scholars of a wide range of expertise who can help us to understand better what tools do scholars need and what are the research questions they expect corpora to answer.

The first session is intended to focus on more general questions of corpus creation, while **the second**, following closely the IMC 2018 topic, aims at showing how the corpora can be practically used in Medieval Latin research.

Session 1: What Corpus for Medieval Studies?

The newly launched project ANR VELUM (Towards Innovative Ways of Visualising, Exploring and Linking Resources for Medieval Latin) aims at constructing a Medieval Latin corpus that would prove useful for the scholarly community. It needs to reflect the variety of medieval writing as well as to provide scholars with a set of analytical tools. Although the term *corpus* is often attributed to haphazard collections of texts, the idea is to build a resource whose architecture would represent in a balanced way proportions of texts as produced during the Middle Ages. *Facile dictu...*

Suggestions: 1) Representativeness: What does it mean for a historical corpus to be representative? How can variety of medieval texts be represented in a limited corpus? What variety of Latin should be represented (time, region, genre *etc.*)? How to assess the importance of a text (manuscript tradition *etc.*)? 2) Texts: What editions to use and how to handle copyright? One corpus for all needs or multiple specialized corpora? 3) What query tools for historical research? What data display and visualisation methods? How to link corpora with other resources (linked data)?

Session 2: Vocabulary of Memory

Vocabulary studies have greatly benefited from the digital turn. As large corpora are now available, the analysis of thousands of occurrences have become easier than ever. Furthermore, metadata that the texts come equipped with facilitate comparative study of genres, single authors *etc.* Methods of distant reading and automatic exploratory analysis are slowly becoming a part of standard equipment of medievalist working with linguistic sources. The session aims at showing how large corpora may help Latinists, historians or philosophers to understand better how the words were used and the concepts verbalized in Medieval Latin. In accordance with the IMC 2018 topic we will focus on the vocabulary of memory.

Suggestions: 1) *Memoria & co.*: What words for memory and similar concepts existed in the Medieval Latin? How did the lexical field evolve through the ages? 2) Does the memory-related vocabulary change with genre, text function or audience it is addressed to? 3) Metaphor: What metaphors was the concept of memory comprehended through? 4) How to automatically retrieve memory-related words from a corpus? How to identify main usage patterns (collocational analysis *etc.*)? How to identify texts concerned with memory (*keywords, topic modelling etc.*) ?

Who?

We are looking for contributions from linguists, historians, digital humanists and other medievalists, both established and early career scholars.

How?

Papers may be delivered in any of the languages of the IMC (English, French, German, Italian, Spanish). Proposals should be addressed **by 15 September 2017** to both bruno.bon@irht.cnrs.fr and krzysztof.nowak@ijp.pan.pl.

Proposal should include:

- author's name and institution;
- title;
- abstract of 150-200 words with max. 2-3 references.

The organizers cannot contribute to the conference and other fees of the session participants.

Should you have any questions, do not hesitate to contact us.

¹ **Bruno Bon** is vice-director of the Institut de Recherche et d'Histoire des Textes (CNRS, Paris), and chief-redactor of the *Novum Glossarium Mediae Latinitatis* (UAI – Comité Du Cange). He has been coordinating the ANR Omnia project (<http://glossaria.eu>) which resulted in retro-digitisation of dictionaries (Du Cange's *Glossarium* and *NGML*) and development of the Treetagger parameters for Medieval Latin lemmatisation, and now coordinates the new ANR Velum project. Bruno's research focuses mainly on historical semantics and textual statistics.

WWW: <https://www.irht.cnrs.fr/annuaire/bon-bruno>

² **Krzysztof Nowak** (PhD) is a Medieval Latin linguist and lexicographer. He is a member of team of the Dictionary of Polish Medieval Latin. He is responsible for the conception, coordination and programming of the eDictionary (<http://scriptores.pl/elexicon/en>) and the eCorpus (<http://scriptores.pl/en/efontes/>) of the Polish Medieval Latin. His recent research focuses on Latin corpus linguistics and distributional semantics.

WWW: <https://ijp-pan.krakow.pl/nasi-pracownicy/doktorzy/krzysztof-nowak>

International Medieval Congress, Leeds, 2-5 Juillet 2018

Appel à communications: *Per corpora...* Latin médiéval et Corpus textuels (2 sessions)

Session 1: Quel corpus textuel pour les études médiévales ?

Session 2: Le vocabulaire de la mémoire

Organisation

Bruno Bon (*Dictionnaire du latin médiéval*, Institut de recherche et d'histoire de textes, CNRS)¹

Krzysztof Nowak (*Dictionnaire du latin médiéval*, Institut de la Langue Polonaise, PAN)²

Motivation

Après une première expérience en 2016, nous souhaitons proposer deux sessions pour l'édition IMC 2018, sur la construction et l'utilisation des corpus textuels électroniques dans les études médiévales, en associant des utilisateurs et créateurs confirmés d'outils numériques à ceux qui font leurs premiers pas dans la recherche sur les corpus. L'IMC offre une belle occasion à ce genre de discussion, car l'événement rassemble des spécialistes de nombreuses disciplines, aptes à signaler les outils qui leur sont nécessaires, et à préciser les questions qu'ils posent à leurs corpus.

La 1^{ère} session traitera des questions générales sur la création de corpus textuels, pendant que **la seconde**, suivant le thème annuel de l'IMC, étudiera des cas d'utilisation pratique des corpus textuels dans la recherche en latin médiéval.

Session 1 : Quel corpus textuel pour les études médiévales ?

Le tout récent projet ANR VELUM (Visualisation, Exploration et Liaison de ressources innovantes pour le latin médiéval) cherche à construire un corpus de latin médiéval le plus utile possible à la communauté scientifique, représentatif de la richesse de l'écriture médiolatine, en l'accompagnant d'outils d'analyse approfondie. Souvent conçu de manière assez lâche, le « corpus textuel » est ici compris de façon rigoureuse, son architecture devant correspondre au moins généralement aux proportions des genres dans la production écrite contemporaine. *Facile dictu...*

Suggestions : 1) Représentativité : Qu'est-ce qu'un corpus textuel historiquement représentatif ? Comment représenter la richesse de la production écrite dans un corpus restreint ? De quel(s) latin(s) s'agit-il (période, région, genre, etc.) ? Comment évaluer l'importance relative des textes médiévaux (tradition manuscrite, etc.) ? 2) Textes : Quelles sources utiliser, dans quelles éditions, avec quels droits ? Un même et unique corpus peut-il répondre à tous les besoins ? 3) Quels outils d'interrogation pour la recherche historique ? Quel affichage des données textuelles ? Quelle visualisation des textes, pour quel usage ? Quel lien avec les autres outils (*linked data*) ?

Session 2: Vocabulaire de la mémoire

Les études du vocabulaire ont largement bénéficié du tournant numérique. Les grands corpus permettent une analyse empirique de milliers d'occurrences, largement libérée de la subjectivité et des biais dont les études lexicales ont longtemps souffert. Les métadonnées rendent possibles l'interrogation simultanée d'un corpus et de ses parties, et les études comparatives de genre, d'auteur ou de périodes. Avec les méthodes automatiques (*distant reading*), on assiste à un changement radical du champ exploratoire des médiévistes. La session cherchera à montrer comment ces corpus peuvent aider le latiniste, l'historien ou le philosophe à mieux comprendre l'usage des mots et la

verbalisation des concepts en latin médiéval. En accord avec le thème de l'IMC 2018, il s'agira ici du vocabulaire de la mémoire.

Suggestions : 1) *Memoria & co.* : Quels mots pour la mémoire et les concepts voisins ? Comment le champ lexical de la mémoire a-t-il évolué à travers les âges ? 2) Fonctions : Comment et pourquoi le vocabulaire lié à la mémoire varie-t-il selon le genre du texte, sa fonction ou son public ? 3) Métaphorisation : Par quelles métaphores le concept de mémoire est-il saisi ? 4) Comment retrouver automatiquement les mots liés à la mémoire ? Comment identifier les principaux modèles d'usage (analyse des cooccurrences, etc.) ? Comment repérer les textes liés à la mémoire (mots-clefs, *topic modelling*, etc.) ?

Qui ?

Les sessions sont ouvertes à tous, linguistes, historiens, humanistes numériques, etc, scientifiques expérimentés ou jeunes chercheurs.

Comment ?

La communication peut se faire dans toutes les langues du congrès (allemand, anglais, français, espagnol, italien). La proposition, qui doit être envoyée au plus tard le **15 septembre 2017** conjointement à bruno.bon@irht.cnrs.fr et krzysztof.nowak@ijp.pan.pl, contiendra :

- le nom et l'institution de l'auteur ;
- le titre de la contribution ;
- un résumé de 150-200 mots ;
- 2-3 références bibliographiques.

Les organisateurs ne peuvent malheureusement pas participer aux frais des participants.

Pour toute question, n'hésitez pas à nous contacter.

¹ **Bruno Bon** est sous-directeur de l'Institut de Recherche et d'Histoire des Textes (CNRS, Paris), et rédacteur du *Novum Glossarium Mediae Latinitatis* (UAI). Il a coordonné le projet ANR Omnia (<http://glossaria.eu>), qui a abouti à la rétrodigitalisation de dictionnaires (*Glossarium* de Du Cange et *NGML*) et au développement de paramètres de lemmatisation du latin médiéval pour Treetagger. Porteur du nouveau projet ANR Velum, il s'occupe actuellement de sémantique historique et de statistique textuelle.

WWW: <https://www.irht.cnrs.fr/annuaire/bon-bruno>.

² **Krzysztof Nowak** est un linguiste et lexicographe du latin médiéval. Membre de l'équipe de rédaction du *Lexicon Medii et Infimae Latinitatis Polonorum* (IJP-PAN), il est responsable de la conception, de la coordination et de la programmation du dictionnaire électronique (<http://scriptores.pl/lexicon/en>) et du corpus électronique (<http://scriptores.pl/en/efontes/>) du latin médiéval polonais. Ses recherches récentes concernent la linguistique de corpus latine et la sémantique distributionnelle.

WWW: <https://ijp-pan.krakow.pl/nasi-pracownicy/doktorzy/krzysztof-nowak>.