

Les coordinateurs
Jean-Gabriel Ganascia, Francesca Frontini

Déclaration d'intention avant le 15 avril 2019.

Site: tal-60-3.sciencesconf.org

APPEL À PROPOSITION

Les humanités numériques (HN) recouvrent aujourd'hui un champ en rapide expansion; si leurs frontières sont à la fois difficiles à identifier et en constante évolution (Dacos and Mounier, 2015; Terras et al., 2013; Ganascia, 2015), leur impact sur les humanités, c'est-à-dire sur les disciplines qui étudient les cultures humaines et/ou les œuvres humaines, ne doit pas être sous-estimé. En effet, la facilité d'accès aux ressources numériques, et en particulier la numérisation des contenus et la façon dont les ordinateurs les traitent, transforme les humanités et ouvre la voie à l'émergence de nouvelles pratiques scientifiques. Étant donné que bon nombre de ces contenus, que ce soit en littérature, en philosophie, en archéologie ou en histoire, sont donnés sous forme textuelle, les techniques du traitement automatique du langage naturel (TAL) sont potentiellement d'un grand profit pour les humanités numériques. Les HN et la recherche actuelle en TAL sont toutes deux issues d'une tradition commune, celle de « l'informatique littéraire et linguistique » (*"Literary and Linguistic Computing"*) (Hockey, 2004). En effet, nombre de chercheurs font souvent remonter les origines des HN au projet *Index Thomisticus* de Roberto Busa qui visait, dès 1949, à créer automatiquement un index de la *Somme théologique* de Thomas d'Aquin avec des ordinateurs. Et, aujourd'hui, le domaine que l'on peut qualifier d'« humanités numériques textuelles » constitue encore un vaste sous-domaine des HN.

Cependant, alors que la recherche actuelle en TAL s'articule généralement autour de tâches bien identifiées et plus ou moins complexes (comme l'étiquetage syntaxique, la lemmatisation, la racinisation (*stemming*), la reconnaissance d'entités nommées, l'analyse syntaxique, l'extraction d'informations, les systèmes question-réponse, le résumé de texte, etc.), les HN utilisent des techniques et des méthodes de TAL comme outils, et les utilisent dans des scénarios de recherche complexes, allant de l'acquisition à l'annotation et l'analyse de textes, et pouvant aussi bien porter sur des collections de textes non structurées, que sur des éditions numériques hautement encodées. En conséquence, alors que les progrès en TAL sont supposés avoir des implications positives pour les recherches dans le secteur des humanités, les défis ultimes dans la perspective des HN ne visent pas uniquement à améliorer les performances des outils de TAL en eux-mêmes, mais leur utilisation en vue d'une recherche innovante qui pourrait vraiment faire avancer la connaissance disciplinaire dans les différents champs des humanités. Au-delà, la taille des corpus peut varier considérablement dans les HN, depuis de grandes bibliothèques de centaines de milliers d'ouvrages numérisées — avec malheureusement de trop fréquentes erreurs — à des petits ensembles de dizaines ou de centaines de livres.

À ces différences de finalité, s'ajoutent la très grande variété et complexité des textes traités. Bien que la recherche en TAL n'ignore pas la nécessité d'adapter ses outils et méthodes à des textes d'époques, de registres ou de genres différents, la diversité des types de textes

communément traités par les HN constitue souvent, par sa nature, un défi supplémentaire pour les outils et algorithmes courants. En particulier, les documents historiques consignés dans des variantes linguistiques anciennes ou les textes littéraires peuvent poser des problèmes tant d'un point de vue linguistique que pour la complexité de leur contenu.

Malgré ou plutôt du fait des problèmes susmentionnés, les applications des HN peuvent se présenter elles-mêmes comme un banc d'essai idéal pour évaluer les dernières avancées dans le TAL.

Ce numéro spécial de la revue TAL sera consacré à la collecte de contributions originales à la croisée des chemins entre la HN et la TAL, avec un accent particulier sur les projets dans lesquels les outils de la TAL sont développés et/ou appliqués pour annoter, traiter et étudier le contenu textuel pour les différentes disciplines des humanités.

Les disciplines couvertes comprendront tous les champs des humanités, depuis la littérature et la philosophie, jusqu'à l'anthropologie et l'histoire. Tous les aspects et niveaux d'analyse du traitement de texte écrit peuvent être concernés, tels que :

- la création de corpus, la numérisation, la transcription
- l'enrichissement et l'annotation automatiques
- l'interrogation et l'exploration de corpus
- l'analyse automatique du texte

Les contributions peuvent concerner les domaines suivants (liste non exhaustive) :

- alignement de texte monolingue ou multilingue
- identification des similitudes de textes, attribution de la paternité d'une œuvre à un auteur, catégorisation de textes
- repérage de références à des œuvres, des individus ou des personnages de fiction
- extraction et annotation des thèmes et des sujets
- extraction de motifs linguistiques récurrents à des fins d'analyse linguistique et stylistique
- détection d'emprunts ou de réutilisations
- adaptation des outils du TAL aux textes et langues historiques
- extraction automatique des connaissances dans le but de créer des ontologies de domaines dans n'importe quel champ des humanités
- outils pour la génétique textuelle
- exploration de grands corpus textuels dans le but d'explorer l'intertextualité ou les variations linguistiques
- exploration de grands corpus textuels pour l'identification d'évolutions culturelles et/ou historiques
- ...

Les articles théoriques et les articles de réflexion seront considérés, à condition qu'ils soient basés sur des recherches et des projets antérieurs des auteurs ou sur des expériences existantes et qu'ils montrent clairement leurs contributions au TAL et aux HN.

RÉFÉRENCES

- Dacos, M. and Mounier, P. (2015). *Humanités Numériques : État Des Lieux et Positionnement de La Recherche Française Dans Le Contexte International*. Research Report Institut français <https://hal.archives-ouvertes.fr/hal-01228945> (accessed 15 January 2019).
- Ganascia J.-G., *The Logic of the Big Data Turn in Digital Literary Studies*, *Frontiers in Digital Humanities*, vol. 2, 2015, number 7, <https://www.frontiersin.org/articles/10.3389/fdigh.2015.00007/full>, ISSN 2297-2668
- Hockey, S. (2004). The History of Humanities Computing. In Schreibman, S., Siemens, R. and Unsworth, J. (eds), *A Companion to Digital Humanities*. Oxford: Blackwell.
- Terras, M., Vanhoutte, E. and Nyhan, J. (2013). *Defining Digital Humanities: A Reader*. London/New York: Routledge.

COORDINATION

Francesca Frontini - Université Paul-Valéry Montpellier 3
Jean-Gabriel Ganascia – Sorbonne Université

CALENDRIER

- diffusion de l'appel à communication : fin janvier 2019
- déclaration d'intention : 15 avril 2019
- Date limite de soumission : 15 mai 2019
- Notification aux auteurs, premières relectures : 15 juillet 2019
- Notification aux auteurs, secondes relectures : 15 octobre 2019^[1]_{SEP}
- Version finale : 30 novembre 2019
- Publication : début 2020

LANGUE

Les articles sont écrits en français ou en anglais. Les soumissions en anglais ne sont acceptées qu'en cas de présence d'au moins un auteur non francophone.

LA REVUE

La revue TAL (Traitement Automatique des Langues) est une revue internationale éditée depuis 1960 par l'ATALA (Association pour le Traitement Automatique des Langues) avec le concours du CNRS. Elle est maintenant publiée en format électronique, avec accès gratuit immédiate aux articles publiés, et impression annuelle à la demande.

CONTACT

tal-60-3@sciencesconf.org

INSTRUCTIONS

Les chercheurs ayant l'intention de soumettre une contribution sont invités à faire une déclaration d'intention en envoyant un mail aux coordinateurs en indiquant le titre prévisionnel et un paragraphe de description.

La soumission de l'article final se fera à travers la plateforme <scienceconf.org>. Ceux qui ne disposent pas d'un compte, sont invités à s'inscrire sur le site <<http://www.sciencesconf.org>> (en haut à gauche, "créer un compte"), puis revenir sur la page <

3.sciencesconf.org>, se connecter et effectuer le dépôt en cliquant sur le menu "Soumission d'un article" (format PDF).

Chaque soumission sera relue par un membre du CR de la revue et deux membres du comité scientifique spécifique de ce numéro spécial.

Les articles doivent faire entre 20 et 25 pages. Les auteurs souhaitant obtenir une dérogation pour la longueur doivent contacter les coordinateurs. La revue TAL a un processus de relecture en double-aveugle. Merci d'anonymiser votre article et le nom du fichier. Les feuilles de style sont disponibles en ligne sur le site de la revue (<http://www.atala.org/content/instructions-aux-auteurs-feuilles-de-style-0>)