
Une ontologie pour Henri et ses amis

Pierre Willaime*¹

¹Archives Henri-Poincaré - Philosophie et Recherches sur les Sciences et les Technologies – université de Strasbourg, Université de Lorraine, Centre National de la Recherche Scientifique : UMR7117 – France

Résumé

Cette communication se propose de présenter méthodologiquement l'édition numérique d'un corpus d'auteur, celui de la correspondance d'Henri Poincaré (1854-1912, ~2100 lettres, henripoincare.fr). Le traitement et l'analyse du corpus posent des questions plus générales, propres aux technologies utilisées (celles du Web sémantique) et à leur applicabilité en humanités numériques. Ce projet est porté collectivement par les Archives Henri-Poincaré (UMR7117 - CNRS, Universités de Lorraine et de Strasbourg).

Le Web sémantique peut être vu comme particulièrement adapté à la description d'un corpus dans le cadre d'un projet d'humanités numériques. En effet, cette extension du Web permet de modéliser les entités, relations, concepts, ou autres métadonnées pouvant décrire les documents d'archives, les acteurs historiques en question et leur contexte. Cette structuration, permise par les technologies standardisées (W3C) constituant le Web sémantique (Berners-Lee et al, 2001) telles que RDF (Resource Description Framework) et OWL (Web Ontology Language), évite l'écueil courant des bases de données relationnelles, adaptées à un projet mais peu interopérables. Plus encore, la "sémantisation" d'un corpus ouvre la voie à son traitement automatique par des méthodes computationnelles.

Travail sur le long cours (depuis 1999), l'édition de ce corpus a été portée sous Omeka S en 2018. Ce CMS a été privilégié à la fois pour sa facilité d'édition, qui permet à tout membre du projet d'avoir un pouvoir d'action, et pour ses fonctionnalités liées au web sémantique.

Omeka S permet facilement de décrire chaque contenu à l'aide de schémas de métadonnées standardisés (Dublin Core dans sa version "DCMI Metadata Terms", Bibliographical ontology, BIO, Relationship, etc.). Cet effort en vue de l'interopérabilité se heurte parfois aux objectifs de recherche, nécessitant souvent des propriétés ad hoc. Les ontologies existantes n'apportent pas l'expressivité d'un encodage en XML-TEI par exemple. Pour ne pas choisir entre d'un côté l'interopérabilité et le respect des standards, et de l'autre une description scientifique fine et adaptée au corpus, nous avons développé une ontologie spécialisée pour l'histoire des sciences et l'édition de correspondances scientifiques. Cette ontologie, nommé AHPo, s'aligne sur les schémas de métadonnées préexistants lorsque cela est possible.

Cette présentation présentera l'ontologie AHPo à l'aide de schémas permettant de comprendre sa structuration. Cette ontologie, et le choix du Web sémantique en général, permettent la mise en place d'outils pour la recherche. Nous en détaillerons deux en présentant les capacités d'interrogation permises par le langage SPARQL et la visualisation du réseau de personnes gravitant autour d'Henri Poincaré. Le langage SPARQL permet de formuler

*Intervenant

des requêtes complexes pour exploiter les liens entre les éléments d'un graphe de données. Cela ouvre de nouvelles possibilités pour le chercheur en termes de recouplement et de fouille fine du corpus. L'outil de visualisation du réseau de personnes permet de son côté une représentation dynamique et temporelle du contexte social d'écriture des lettres. Il est basé sur l'ontologie et sa description fine des relations entre entités.

De manière plus générale, cette communication souhaite mettre en avant les possibilités techniques d'exploitation de corpus permises par le Web sémantique. Si elles ne permettent pas d'atteindre une granularité aussi fine qu'un balisage XML-TEI détaillé, les technologies sémantiques mettent cependant l'accent sur le corpus vu comme un ensemble en décrivant précisément les relations entre entités et en allant plus loin qu'une simple édition numérique. Les deux technologies peuvent conduire à des visions complémentaires d'un même corpus. Il serait intéressant de coupler ces usages, présentés souvent comme deux approches incompatibles et comme deux directions pour un projet d'humanités numériques.

Mots-Clés: Web sémantique, ontologie, visualisation de réseau, Henri Poincaré, Omeka S