# Poster proposal

# TEL : An open access corpus available for research

Agnès Magron

Centre pour la Communication Scientifique Directe, Villeurbanne

TEL (Theses-en-Ligne) (<https://tel.archives-ouvertes.fr>) is a french open repository dedicated to theses self-archiving. It has been created in 2001 and is a particuliar instance of HAL. It is multidisciplinary, and available to any user who has defended a thesis, whatever his university. Moreover, with its partnership with ABES, it occupies a central position in the dissemination of theses. Indeed, TEL is supplied by the STAR application, which allows academic institutions to submit defended theses under a digital format.

In January 2016, it contains the full text of more than 53000 theses, the older one dating back to 1907. It forms a remarkable corpus available to researchers interested in data from these theses, whether they're interested in a specific domain, or for a comparison purpose. Since TEL is an HAL portal, the data can also be observed in the author's other publications, if they've been submitted to HAL. Metadata describing the theses (author, universities, jury members…) also form a corpus that can be used for any study on doctoral research in France. Indeed, API allow to search through them with a high level of accuracy. They can be used for forming a corpus, producing statistics, cross-check data, and compare them to other corpus data.

The poster presents the corpus with some statistical data.

Poster : TEL : un corpus en libre accès pour la recherche

TEL (Theses-en-Ligne) (<https://tel.archives-ouvertes.fr>) est une archive ouverte française dédiée à l’auto-archivage des thèses. Elle a été créée en 2001 et est un environnement spécifique de HAL. Elle a la particularité d’être multidisciplinaire et d’être à la disposition de tout usager ayant soutenu sa thèse, quelle que soit son université. De plus, par son partenariat avec l’ABES, elle occupe une place particulière dans la dissémination des thèses. TEL est en effet alimenté par l’application STAR, qui permet aux établissements de soutenance de déposer les thèses soutenues sous forme électronique.

En janvier 2016, elle contient le texte intégral de plus de 53 000 thèses, la plus ancienne datant de 1907. Elle constitue un corpus remarquable à la disposition des chercheurs s’intéressant aux données présentes dans les thèses, qu’ils s’intéressent à une discipline en particulier ou pour établir des comparaisons. Parce que TEL est un portail de HAL, *le devenir de ces données* peut aussi être observé dans les publications du même auteur si elles sont déposées elles aussi dans HAL.

Les métadonnées décrivant les thèses (auteur, universités, composition des jurys, etc) constituent elles-mêmes un corpus sur lequel s’appuyer pour mener toute étude portant sur la recherche doctorale en France. Elles sont en effet interrogeables de manière très fine via les API. Celles-ci permettent de se constituer un corpus, d’établir des statistiques, de croiser des données, de les confronter avec celles d’autres corpus.

Le poster présente quelques statistiques sur la composition du corpus TEL et les outils mis à disposition.

Bibliographie :

O. Cian et M. Giloux, STAR et le dépôt électronique des thèses, Arabesques, 2015, n°78, p.8-9

J. Schöpfel, H. Prost, C. Malleret. Making data in PhD dissertations reusable for research. *8th Conference on Grey Literature and Repositories*, Oct 2015, Prague, Czech Republic. 2015, [<http://nrgl.techlib.cz/index.php/Proceedings\_2015>](http://nrgl.techlib.cz/index.php/Proceedings_2015). [<hal-01248979>](http://hal.univ-lille3.fr/hal-01248979)