# Avis de Soutenance

## Madame Yutong FEI

## Sciences de l'information et de la communication

Soutiendra publiquement ses travaux de thèse intitulés
*Étude et classification des contextes de citation pour une construction d'indicateurs relationnels et sémantiques*

Travaux dirigés par Madame Chérifa BOUKACEM - ZEGHMOURI

Soutenance prévue le ***mercredi 10 juin 2026*** à 14h30
Lieu : Salle Fontannes (RdC) – Bâtiment Darwin, 3-9 rue Raphaël Dubois, 69100 Villeurbanne

**Composition du jury proposé**

| | | | |
|---|---|---|---|
| Mme Chérifa BOUKACEM - ZEGHMOURI | Professeure des universités | Lyon 1 Université | Directrice de thèse |
| M. Adrian STAII | Professeur des universités | Université Jean Moulin Lyon 3 | Examinateur |
| Mme Nathalie PINEDE | Professeure des universités | Université Bordeaux Montaigne | Rapporteure |
| M. Marc BERTIN | Maître de conférences | Université Marie et Louis Pasteur Montbéliard | Co-encadrant de thèse |
| Mme Béatrice MILARD | Professeure des universités | Université Toulouse – Jean Jaurès | Examinatrice |
| M. Thierry LAFOUGE | Professeur émérite | Lyon 1 Université | Examinateur |
| M. Philippe MONGEON | Professeur adjoint | Université Dalhousie - Halifax (Canada) | Rapporteur |
| M. Abdelghani MADDI | Ingénieur de recherche | Sorbonne Université | Examinateur |

**Mots-clés :** Contexte de citation, Libre Accès, Édition scientifique, Classification des citations, Information scientifique et technique, Bibliométrie

**Résumé :**

Depuis la première proposition d'une typologie de catégories sémantiques citationnelles en 1965 par Eugene Garfield, pour mieux comprendre les raisons et les fonctions de l'acte de citer, ce sujet de recherche n'a cessé de susciter l'intérêt de plusieurs disciplines, telles que la sociologie des sciences, la linguistique, la bibliométrie et l'informatique. Dans cette perspective, les travaux empiriques ont évolué, passant des entretiens sur les motivations de citation à la classification sémantique des contextes de citation, c'est-à-dire les passages textuels entourant une référence dans un article scientifique. Avec le développement des approches informatiques, cette classification tend aujourd'hui à s'automatiser, donnant lieu à des services à valeur ajoutée intégrés dans les systèmes de recommandation de citations destinés à améliorer la recherche bibliographique. Or, comprendre

la citation suppose d'abord de préciser ce que recouvre la sémantique citationnelle et d'en saisir la genèse. Celle-ci ne se réduit pas à une analyse linguistique, mais reflète les stratégies scientifiques, sociales, rhétoriques, et discursives des auteurs. Les formes de citation résultent ainsi de multiples facteurs, tels que les spécificités disciplinaires, les pratiques de recherche ou encore l'évolution des normes éditoriales au fil du temps. Réduire les contextes de citation à de simples ressources d'entraînement pour des applications industrielles appauvrit la compréhension de la citation, en négligeant les tensions qui structurent cette complexité. Cet enjeu est au cœur de cette thèse, qui construit un corpus de contextes de citation prenant en compte les facteurs influençant la production du discours citationnel, notamment la diversité disciplinaire et éditoriale, afin de proposer une typologie des différentes manières de citer. Le corpus réunit 8 358 contextes issus de 157 revues internationales, publiées entre 2020 et 2023, période marquée par une tendance à la standardisation des politiques éditoriales autour des valeurs d'ouverture et de transparence de la recherche. En complément de cette analyse discursive, une étude exploratoire fondée sur des entretiens semi-directifs avec trois acteurs phares du marché de la citation a été réalisée pour interroger l'usage industriel des contextes de citation. Les 50 catégories de citation dégagées dans cette thèse révèlent une grande diversité sémantique, façonnée par la tension entre les caractéristiques disciplinaires du discours citationnel et la standardisation éditoriale. Les différentes manières de construire le savoir propre à chaque discipline se reflètent dans cette diversité, telle qu'elle s'exprime à travers la dimension discursive de la citation, car la construction du savoir et celle du discours sont indissociables, particulièrement dans les SHS. Par ailleurs, la structuration de cette diversité discursive par les lignes éditoriales constitue un enjeu industriel, en lien avec la future exploitation des contextes de citation dans les bases de données. Les résultats de cette recherche doctorale montrent les effets des cadres politiques et institutionnels de la recherche sur les cultures discursives de la citation. De même, ils montrent de quelle manière l'implémentation des contextes de citation ouvre la voie à de nouvelles formes d'exploitation de la publication scientifique ouverte, qui offre aux acteurs du marché des opportunités renouvelées de marchandisation.

**Summary:**

Since Eugene Garfield first proposed a typology of citation semantics in 1965 with the objective of a better understanding of the reasons and functions behind the act of citing in scientific articles, this research subject has received sustained attention from multiple disciplines, including the sociology of science, linguistics, bibliometrics, and computer science. In this regard, empirical studies have evolved from interviews exploring authors' motivations for citing to classifying citation contexts, i.e., the textual blocks surrounding a cited reference in a scientific article. In light of advances in computational approaches, citation context classification is increasingly automated, leading to data-driven information services integrated into citation recommendation systems aimed at improving bibliographic search. However, understanding citation behavior primarily requires clarifying what is meant by citation semantics and why different semantics underlie citation behaviors. This cannot be reduced to a purely linguistic analysis, rather, it reflects the scientific, social, rhetorical, and discursive strategies of authors. Citation behavior thus results from multiple factors, including disciplinary specificities, research practices, and the evolution of editorial norms over time. Reducing citation contexts to simple training data for industrial applications oversimplifies our understanding of citation behavior, by overlooking the multiple dimensions of tension that shape its inherent complexity. Building on this issue, we constructed a corpus of citation contexts that accounts for factors influencing the production of citation discourse, particularly disciplinary and editorial diversity, in order to propose a typology of different citation behaviors. The corpus comprises 8,358 contexts drawn from 157 international journals published between 2020 and 2023, a period characterized by a trend toward the standardization of author guidelines emphasizing openness and transparency in research. In addition to this discursive analysis of citation contexts, we conducted

an exploratory study based on semi-structured interviews with three key stakeholders in the citation market to further investigate the industrial applications of citation contexts. We identified 50 categories of citation behavior in our corpus, which revealed considerable semantic diversity, shaped by the tension between the disciplinary characteristics of citation discourse and editorial standardization. Our findings suggest that the different ways knowledge is constructed within each discipline are reflected in this diversity, as expressed through the discursive dimension of citation, since knowledge construction and discourse construction are inseparable, particularly in the social sciences and humanities. Moreover, the standardization of this scientific discourse diversity through author guidelines also provides the citation context resources for integration into scientific databases. This doctoral research thus highlights the effects of political and institutional research frameworks on the discursive cultures of citation. Finally, we show how the implementation of citation contexts in the citation market paves the way for new forms of scientific publication exploitation, offering actors new opportunities to commercialize scientific information based on full texts and their semantics.