

# Extraction et reconstitution de relation n-Aires issues d'articles scientifiques en domaine expérimental guidées par une Ressource Termino-Ontologique

Martin Lentschat<sup>\*,\*\*</sup>, Patrice Buche<sup>\*</sup>, Juliette Dibie<sup>\*\*\*</sup>, Mathieu Roche<sup>\*\*</sup>

<sup>\*</sup>IATE, Univ Montpellier, INRAE, Institut Agro, Montpellier 34060, France

<sup>\*\*</sup>TETIS, Univ. Montpellier, AgroParisTech, CIRAD, CNRS, INRAE, France

<sup>\*\*\*</sup>Univ. Paris-Saclay, INRAE, AgroParisTech, UMR MIA-Paris, France

## 1 Contexte de contribution

Le but de ce travail est l'extraction d'informations expérimentales dans le domaine des emballages alimentaires et leur formalisation dans des relations n-Aires. Les difficultés d'extraction de ces relations proviennent de la dispersion des instances d'arguments dans l'ensemble des documents et de la multiplicité des instances à mettre en relation. Nous comparons différentes méthodes s'appuyant sur une nouvelle représentation multi-descripteurs des relations n-Aires partielles présentes dans les tableaux des articles (STaRe - Scientific Table Representation) pouvant être complétées par les instances d'arguments issues du texte.

Notre méthode de reconstruction des relations n-Aires est guidée par une Ressource Termino-Ontologique (RTO) et repose sur des représentations multi-descripteurs des instances d'arguments issues du texte (SciPuRe - Scientific Publication Representation) (Lentschat et al., 2022) et des instances de relations n-Aires partielles issues des tableaux (STaRe - Scientific Table Representation, e.g. Table 1). STaRe comprend plusieurs types de descripteurs : (i) *ontologiques* qui classifient la relation selon le formalisme de la RTO, (ii) *structurels* qui donnent le contexte du tableau et à sa place dans le document et (iii) *lexicaux* qui indique la forme textuelle des instances d'arguments composant la relation.

Les relations obtenues sur la base de ces représentations sont reconstituées selon trois stratégies exploitant les descripteurs de STaRe et de SciPuRe. L'*approche structurelle* exploite la structure des articles. L'*approche fréquentiste* recherche les co-occurrences fréquentes. L'*approche par plongements lexicaux* utilise les scores de similarité donnés par des modèles de langage.

## 2 Résultats et discussions

Dans l'évaluation des approches de complétion des instances de relations partielles, nous considérons deux aspects : le filtrage des instances d'arguments candidates et la sélection des candidats. Nos expérimentations réalisées sur un jeu de données dédié (Lentschat et al., 2021) ont montré que les mesures proposées permettent d'améliorer à la fois la précision et le rappel